STATS 506 Final Report
Rural Nonprofits and Government Funding

Calder Moore (caldermo)

December 8, 2025

Abstract

In this report, I use the The National Survey of Nonprofit Trends and Impacts to explore whether nonprofit organizations having operations in rural areas is associated with a higher proportion of their funding coming from government agencies, controlling for covariates. I use a fractional logistic regression to answer this question. We find that there is a statistically significant and negative relationship between nonprofits having operations in rural areas and the proportion of their funding that comes from government agencies.

1 Introduction

Nonprofit organizations have operations in a wide variety of places, although it is much easier for operations to take root in more urban settings due to the greater availability of location and the higher density of people that the group can serve. Nonprofits also rely on a wide variety of funding sources to keep their operations afloat, one of which is government agencies that may have an interest in supporting their programs. Due to the possible increased barrier to entry in setting up operations in rural areas, I was interested in exploring if there is a positive relationship between nonprofits that have operations in a rural area and a higher proportion of funding coming from government agencies.

2 Data

The data I used is The National Survey of Nonprofit Trends and Impacts. I only used variables for the year 2021. The raw data only included dollar values in absolute terms for the various funding sources, so I created several variables out of each but that measured the proportion of each group's funding that came from each source. The main variable of interest here, which served as the response in my model, was `FundPropGovAg` which was the proportion of funding coming from government agencies. I also included a variable for the number of staff, which I created by summing the other staff count variables. This was to serve as a proxy for size of the organization, which could have had an impact as a potential lurking variable due to larger organizations having an easier time opening rural locations or getting preferential treatment in government grants. Similarly, a number of dummies to indicate the presence of programs aimed at certain groups are included since they too could have a similar impact when it comes to attracting government funds, depending on what sort of programs each group is offering.

The data included many entries of -99 to indicate when a respondent saw the question, but chose not to respond. I made these values `NA` for the finance variables since it doesn't necessarily mean they have no funding from that particular source, but maybe they just didn't care to disclose that information. However for the indicators for urban or rural operations, the variables only started with two levels (-99 and 1) so I changed the -99 entries to 0 as it seemed reasonable that organizations that didn't apply to one would simply leave it blank. For the staffing variables I turned -99 variables into `NA` for similar reason to the finance variables, and the program variables I turned both -99 and 99 values into `NA`. The program variables also had 3 levels once the -99 and 99 values were taken care of: 0 for no program for this group, 1 for this group is a primary group that we work with, and 2 if the group was a secondary group. I didn't think that the distinction between primary and secondary groups was necessary for this analysis, so I made the 2 values into 1 to have a two level structure, just an indicator for if the nonprofit had a program at all for this group or not.

See Figure 1 in the appendix for a table of summary statistics that were used in the analysis.

3 Methodology

Because the response variable is a proportion and included values exactly equal to 0 or 1, I use a quasibinomial (or fractional logistic) model for analysis.

$$\text{FundPropGovAg}_p = \beta_0 + \beta_1 \text{RuralUrban\_1} + \beta_2 \text{TotalStaff} + \sum_{k \in K} \beta_k \text{ProgDem}_k \tag{1}$$

For nonprofit p. K is the set of natural numbers from 1 to 21, excluding 3, 11, 14, 17, and 18 since they were removed after reviewing variance inflation factors to better meet the no multicollinearity assumptions of the model. I am interested in the value of β_1. The other predictors are included to serve as covariates for other features of a nonprofit that could have some impact on

the funding source. The number of staff is intended to be a proxy for the size of an organization, and the program variables are intended to account for the difference in programs for certain groups offered by different nonprofits. See Figure 3 for a histogram of the distribution of the response variable, `FundPropGovAg`. In using many of the `ProgDem` variables as controls for various programs offered, there was a risk of high multicollinearity that could have an impact on the model. After checking variance inflation factors and seeing they were quite high when including all but one of these variables, after only dropping the 5 variables with the highest variance inflation factors, the factors looked much better, all being lower than 10.

4 Results

I find that there is a statistically significant negative relationship between nonprofits having operations in a rural area and the proportion of their funding that comes from government agencies. In particular, because we have a logit link in our quasibinomial regression, we would expect that, on average, a nonprofit having operations in a rural area is associated with $100(e^{-1.237} - 1) = -70.97\%$ of their funding coming from government agencies compared to nonprofits without rural operations, as a proportion of their total funding. This suggests that perhaps the government is more interested in aiding nonprofits with operations elsewhere, or perhaps that rural nonprofits simply are able to adequately fund their programs with income from other sources.

See Figure 2 for a table of the regression coefficients.

5 Conclusions

To conclude, using the quasibinomial regression model shown in equation (1), I found a statistically significant negative association between a nonprofit having operations in a rural area and the proportion of that nonprofit's funding that comes from government agencies.

Using the number of staff here as a proxy for nonprofit size could be a bit dubious, since there are other ways one could determine the "size" of a nonprofit. Expenses or number of people served could also be suitable controls for this. Preserving the two level structure for the program variables could be more interesting as well, to see if there is a difference between nonprofits that focus on certain groups as a primary group vs secondary group that they serve. Running another type of model that can also account for a response on [0,1] would help the analysis by being able to compare them.

Click here for a link to a Github repository which stores the code used for this report. The link is also below in case the embedded link does not work.

https://github.com/moorecal/STATS-506-Final-Project

6 Attribution of Sources

A Pew Research article discussing the `survey` package in R and how to use it with weighted survey data was useful in helping to find functions that can deal with the weights, and how to implement it in my own analysis.

https://www.pewresearch.org/decoded/2019/01/15/a-short-intro-to-linear-regression-analysis-using-survey-data/

ChatGPT was also useful in helping to fix errors that I made during my data cleaning and modeling process, as well as formatting the tables in the appendix to be more readable.

7 Appendix

Figure 1

Statistic	N	Mean	St. Dev.	Min	Max
FundPropGovAg	915	0.365	0.325	0.000	1.000
RuralUrban_1	2,306	0.540	0.498	0	1
TotalStaff	2,306	150.474	632.000	0.000	15,020.000
ProgDem_1	1,973	0.818	0.386	0	1
ProgDem_2	1,906	0.789	0.408	0	1
ProgDem_3	1,964	0.843	0.364	0	1
ProgDem_4	1,863	0.768	0.422	0	1
ProgDem_5	1,585	0.415	0.493	0	1
ProgDem_6	1,774	0.742	0.437	0	1
ProgDem_7	1,835	0.744	0.437	0	1
ProgDem_8	1,739	0.665	0.472	0	1
ProgDem_9	1,624	0.473	0.499	0	1
ProgDem_10	1,779	0.589	0.492	0	1
ProgDem_11	1,791	0.609	0.488	0	1
ProgDem_12	1,678	0.484	0.500	0	1
ProgDem_13	1,691	0.490	0.500	0	1
ProgDem_14	1,585	0.422	0.494	0	1
ProgDem_15	1,708	0.545	0.498	0	1
ProgDem_16	1,766	0.595	0.491	0	1

Figure 2

Variable	Coefficient	SE
(Intercept)	1.205+	0.621
RuralUrban_1	−1.237*	0.478
TotalStaff	−0.002	0.001
ProgDem_1	−0.272	0.759
ProgDem_2	0.673	0.669
ProgDem_4	−0.306	0.469
ProgDem_5	2.123**	0.634
ProgDem_6	0.187	0.597
ProgDem_7	0.690	0.860
ProgDem_8	−2.010*	0.747
ProgDem_9	0.202	0.881
ProgDem_10	−1.188	0.745
ProgDem_12	−0.783	0.821
ProgDem_13	1.837+	0.963
ProgDem_15	0.024	0.591
ProgDem_16	1.220	0.819
ProgDem_19	−1.666+	0.907
ProgDem_20	−0.838	0.559
ProgDem_21	−0.054	0.807

+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001

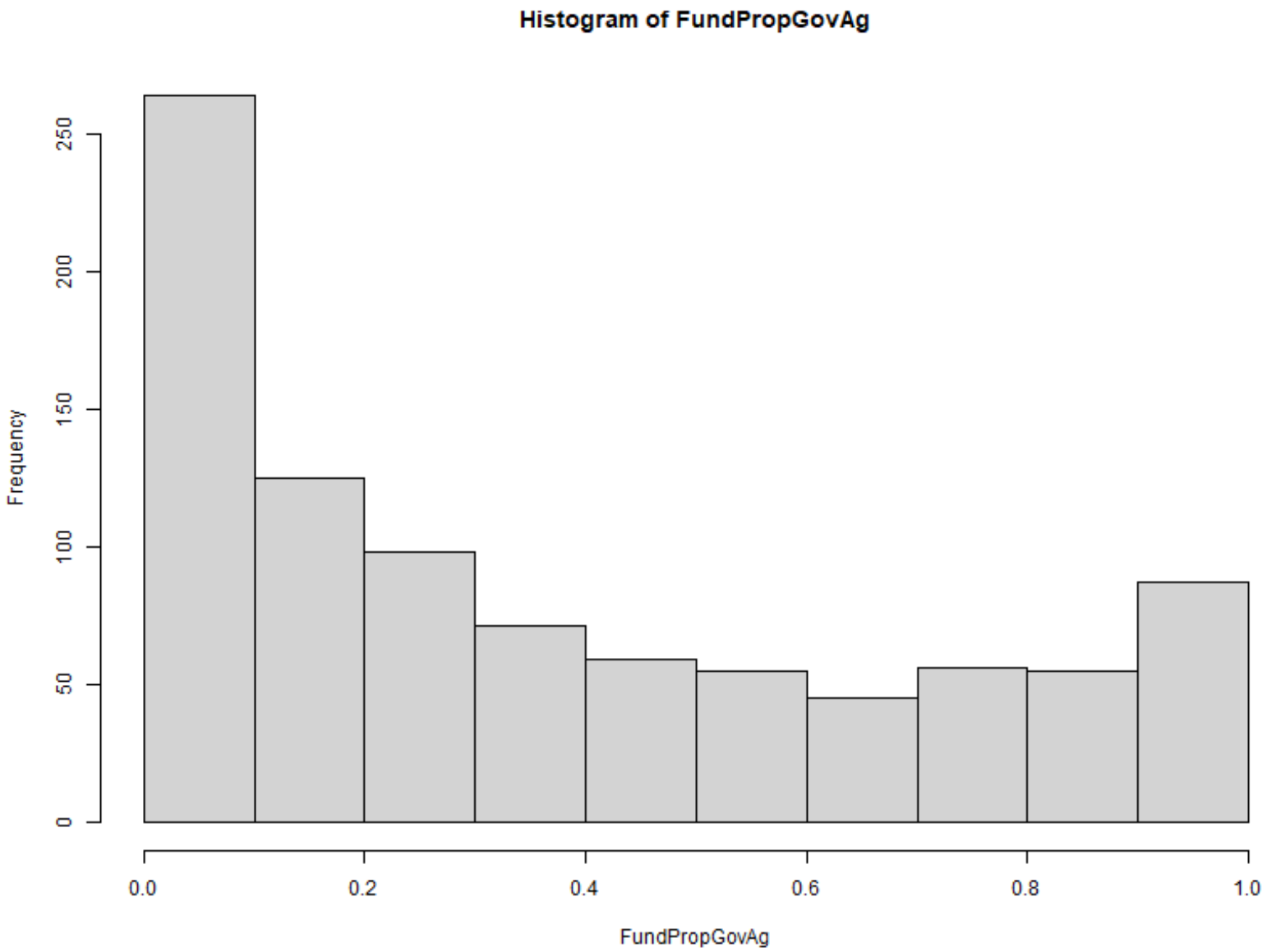

Figure 3: Histogram of Proportion of Funding from Government Agencies