Applied Statistical Analysis II

Replication Study by 23335541 Dan Zhang

Date: April 2024

It's Not Just What You Have, but Who You Know: Networks, Social Proximity to Elites, and Voting in State and Local Elections

- Source:
    - Cambridge University Press, 23 February 2017
    - Doi: 10.1017/S000305541600071X

- **Authors:**
    - Matthew T. Pietryka and Donald A. Debats

“

Individual-level studies of electoral turnout and vote choice have focused largely on personal attributes as explanatory variables. We argue that scholars should also consider the social network in which individuals are embedded, which may influence voting through variation in individuals’ social proximity to elites.

”

(Pietryka and Debats, 2017, p. 360)

Paper Overviews

Question:
How individual electoral voting behavior and vote choice are affected by an individual's social proximity to elites in a social network?

Hypothesis:

Individuals more socially proximate to a city's elites should be more likely to turnout to vote in elections.

Data

- From the 1859 statewide elections in Alexandria, Virginia
- The data set has the ability to identify the network locations of potential voters and a candidate running for local elected office

hhwealth	hhwealthlog	midstatus	highstatus	owner	age	agelog	church	usborn	z1ev	z1elite_avgprox
0.050	−2.975929737	0	0	0	60	4.094345	0	1	4.57110691	0.828745
0.050	−2.975929737	1	0	0	35	3.555348	0	1	−0.22592667	−0.088215
0.050	−2.975929737	1	0	0	45	3.806663	0	1	−0.22599848	−0.350204
0.530	−0.632993340	0	0	0	42	3.737670	0	0	4.57082939	0.828745
0.052	−2.937463284	1	0	0	52	3.951244	1	0	−0.14933088	1.068902
0.021	−3.816712856	1	0	0	43	3.761200	0	1	−0.22235741	0.774164

Alexandria data

Method-interest of variables

- **Explanatory variables:**
- Personal attributes (Control variables)
- An individual's weighted eigenvector network centrality
- An individual's social proximity to elites

- **Outcome variables:**
- An individual's probability of voting

Method-Models

Model approaches：	Logistic regression models are used to predict an Individual’s probability of voting
For each interest outcome, run 3 difference models with difference explanatory variables:	**Model 1**: only include Personal attributes (Control variables)
	Model 2: introduce An individual’s weighted eigenvector network centrality
	Model 3: introduce measure of an individual’s elite proximity(Social proximity to elites)

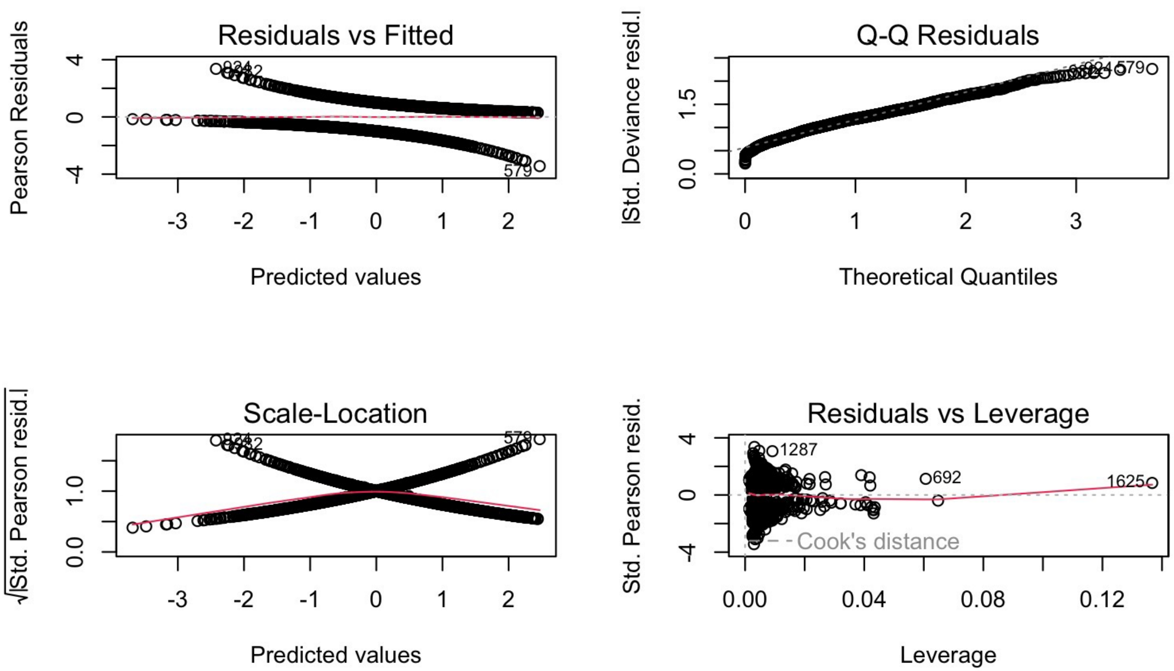
Replication

Original paper: add the status occupation as two separate variables

Replication study: consider an ordered status occupation variable for the models

Replication-Check model assumptions

- First, let's check the assumptions:



Replication-Check model assumptions

- Conduct a Hosmer and Lemeshow goodness fit:

```
        Hosmer and Lemeshow goodness of fit (GOF) test

data:  alexandria_turnout_1$y, fitted(alexandria_turnout_1)
X-squared = 1.5468, df = 8, p-value = 0.9919

        Hosmer and Lemeshow goodness of fit (GOF) test

data:  alexandria_turnout_2$y, fitted(alexandria_turnout_2)
X-squared = 10.424, df = 8, p-value = 0.2365

        Hosmer and Lemeshow goodness of fit (GOF) test

data:  alexandria_turnout_3$y, fitted(alexandria_turnout_3)
X-squared = 6.0518, df = 8, p-value = 0.6414
```

P values greater than 0.05, can not reject null hypothesis.

Replication
-Create ordered status occupation variable

```
# Create 'statuslevel' based on 'midstatus' and 'highstatus'
alexandria_data$statuslevel <- ifelse(alexandria_data$midstatus == 1, "middle status",
                    ifelse(alexandria_data$highstatus == 1, "high status", "non-status"))

# Convert 'statuslevel' to an ordered factor
alexandria_data$statuslevel <- factor(alexandria_data$statuslevel, levels = c("non-status",
"middle status", "high status"), ordered = TRUE)
```

Replication-Apply ordered status variables

```
# Create new control variables
new_controls <- c(
  "hhwealth",    # Household wealth (thousands of dollars)
  "hhwealthlog", # ln(Household wealth)
  "statuslevel", # Status occupation , ordered
  "owner",       # Owns home?
  "age",         # Age (years)
  "agelog",      # ln(Age)
  "church",      # Is church member?
  "usborn"       # Is U.S. born?
) %>%
  paste(collapse = " + ")
```

Replication-Fit new models

```
# MODEL 7
alexandria_turnout_7 <- paste("turnout ~ ",
                 new_controls)  %>% formula()  %>%
 glm(data = alexandria_data,
   family = binomial(link = "logit"))

# MODEL 8
alexandria_turnout_8 <- paste("turnout ~ ",
                 new_controls,
                 "+ z1ev")  %>% formula()  %>%
 glm(data = alexandria_data,
   family = binomial(link = "logit"))

# MODEL 9
alexandria_turnout_9 <- paste("turnout ~ ",
                 new_controls,
                 "+ z1ev + z1elite_avgprox")  %>% formula()  %>%
 glm(data = alexandria_data,
   family = binomial(link = "logit"))
```

Replication- Compare with original models

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-13.500 *	-13.806 *	-14.244 *	-12.983 *	-13.456 *	-13.952 *
	(2.146)	(2.148)	(2.167)	(2.150)	(2.154)	(2.174)
hhwealth	-0.011 *	-0.010 *	-0.010 *	-0.011 *	-0.010 *	-0.010 *
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
hhwealthlog	0.063 *	0.041	0.043 *	0.063 *	0.041	0.043 *
	(0.021)	(0.022)	(0.022)	(0.021)	(0.022)	(0.022)
midstatus	0.687 *	0.417 *	0.290 *			
	(0.103)	(0.119)	(0.123)			
highstatus	0.864 *	0.633 *	0.587 *			
	(0.186)	(0.193)	(0.194)			
owner	0.765 *	0.802 *	0.781 *	0.765 *	0.802 *	0.781 *
	(0.150)	(0.151)	(0.152)	(0.150)	(0.151)	(0.152)
age	-0.091 *	-0.097 *	-0.099 *	-0.091 *	-0.097 *	-0.099 *
	(0.020)	(0.020)	(0.021)	(0.020)	(0.020)	(0.021)
agelog	4.378 *	4.615 *	4.738 *	4.378 *	4.615 *	4.738 *
	(0.808)	(0.810)	(0.817)	(0.808)	(0.810)	(0.817)
church	0.881 *	0.879 *	0.582 *	0.881 *	0.879 *	0.582 *
	(0.106)	(0.107)	(0.125)	(0.106)	(0.107)	(0.125)
usborn	0.723 *	0.606 *	0.543 *	0.723 *	0.606 *	0.543 *
	(0.119)	(0.123)	(0.124)	(0.119)	(0.123)	(0.124)
z1ev		-0.162 *	-0.235 *		-0.162 *	-0.235 *
		(0.038)	(0.041)		(0.038)	(0.041)
z1elite_avgprox			0.557 *			0.557 *
			(0.125)			(0.125)
statuslevel.L				0.611 *	0.448 *	0.415 *
				(0.131)	(0.136)	(0.137)
statuslevel.Q				-0.208 *	-0.082	0.003
				(0.090)	(0.094)	(0.096)
AIC	2697.508	2680.728	2659.522	2697.508	2680.728	2659.522
BIC	2754.543	2743.466	2727.964	2754.543	2743.466	2727.964
Log Likelihood	-1338.754	-1329.364	-1317.761	-1338.754	-1329.364	-1317.761
Deviance	2677.508	2658.728	2635.522	2677.508	2658.728	2635.522
Num. obs.	2216	2216	2216	2216	2216	2216

* $p < 0.05$

Replication-Check model fitting again

```
        Hosmer and Lemeshow goodness of fit (GOF) test

data:  alexandria_turnout_7$y, fitted(alexandria_turnout_7)
X-squared = 1.5468, df = 8, p-value = 0.9919


        Hosmer and Lemeshow goodness of fit (GOF) test

data:  alexandria_turnout_8$y, fitted(alexandria_turnout_8)
X-squared = 10.424, df = 8, p-value = 0.2365


        Hosmer and Lemeshow goodness of fit (GOF) test

data:  alexandria_turnout_9$y, fitted(alexandria_turnout_9)
X-squared = 6.0518, df = 8, p-value = 0.6414
```

```
Analysis of Deviance Table

Model 1: turnout ~ hhwealth + hhwealthlog + midstatus + highstatus + owner +
    age + agelog + church + usborn + z1ev + z1elite_avgprox
Model 2: turnout ~ hhwealth + hhwealthlog + statuslevel + owner + age +
    agelog + church + usborn + z1ev + z1elite_avgprox
  Resid. Df Resid. Dev Df   Deviance Pr(>Chi)
1      2204     2635.5
2      2204     2635.5  0 9.0949e-13
```

No too much difference between the two models, GDF and LRT tests show the evidence.