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

Dataset: From the 1859 statewide elections in Alexandria, Virginia

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

Explanatory variables:

- Personal attributes (Control variables)
  - "hhwealth", # Household wealth (thousands of dollars)" , continuous
  - "hhwealthlog", # ln(Household wealth) , continuous
  - "midstatus", # Mid-status occupation , categorical 0/1
  - "highstatus", # High-status occupation, categorical 0/1
  - "owner", # Owns home? , categorical 0/1
  - "age", # Age (years) , continuous
  - "agelog", # ln(Age) , continuous
  - "church", # Is church member? , categorical 0/1
  - "usborn" # Is U.S. born? , categorical 0/1
- An individual's weighted eigenvector network centrality ("zlev")
- An individual's social proximity to elites ("zlelite_avgprox")

Outcome variables:

- An individual's probability of voting ("turnout")

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)

Method-Results

- The models suggest, holding other variables constant, people who have greater wealth, have higher-status occupations, and own homes are all more likely to vote.

- Model 3 shows a statistically significant negative coefficient associated with network centrality. All else equal, more central citizens were less likely to vote than those more disconnected from the community.

- Model 3 shows a positive and statistically significant coefficient associated with elite proximity, suggesting people more proximate to elites vote at higher rates.

- Conclusion: the effects of centrality and social proximity in social networks on voting are independent and in opposite directions, and therefore cannot substitute for each other. Fowler and Smirnov and Siegel's model shows that the impact of network centrality on voting depends on the overall level of participation in an election.

	Alexandria		
	(1)	(2)	(3)
Household wealth (thousands of dollars)	−0.011*	−0.010*	−0.010*
	(0.004)	(0.004)	(0.004)
ln(household wealth)	0.063*	0.041	0.043*
	(0.021)	(0.022)	(0.022)
Mid-status occupation	0.687*	0.417*	0.290*
	(0.103)	(0.119)	(0.123)
High-status occupation	0.864*	0.633*	0.587*
	(0.186)	(0.193)	(0.194)
Owns home?	0.765*	0.802*	0.781*
	(0.150)	(0.151)	(0.152)
Age (years)	−0.091*	−0.097*	−0.099*
	(0.020)	(0.020)	(0.021)
ln(age)	4.378*	4.615*	4.738*
	(0.808)	(0.810)	(0.817)
Is church member?	0.881*	0.879*	0.582*
	(0.106)	(0.107)	(0.125)
Is African American?	—	—	—
	—	—	—
Is U.S. born?	0.723*	0.606*	0.543*
	(0.119)	(0.123)	(0.124)
Network centrality (*Z* score)	—	−0.162*	−0.235*
	—	(0.038)	(0.041)
Social proximity to elites (*Z* score)	—	—	0.557*
	—	—	(0.125)
Intercept	−13.500*	−13.806*	−14.244*
	(2.146)	(2.148)	(2.167)
N	2216	2216	2216
Log likelihood	−1338.754	−1329.364	−1317.761
AIC	2697.508	2680.728	2659.522

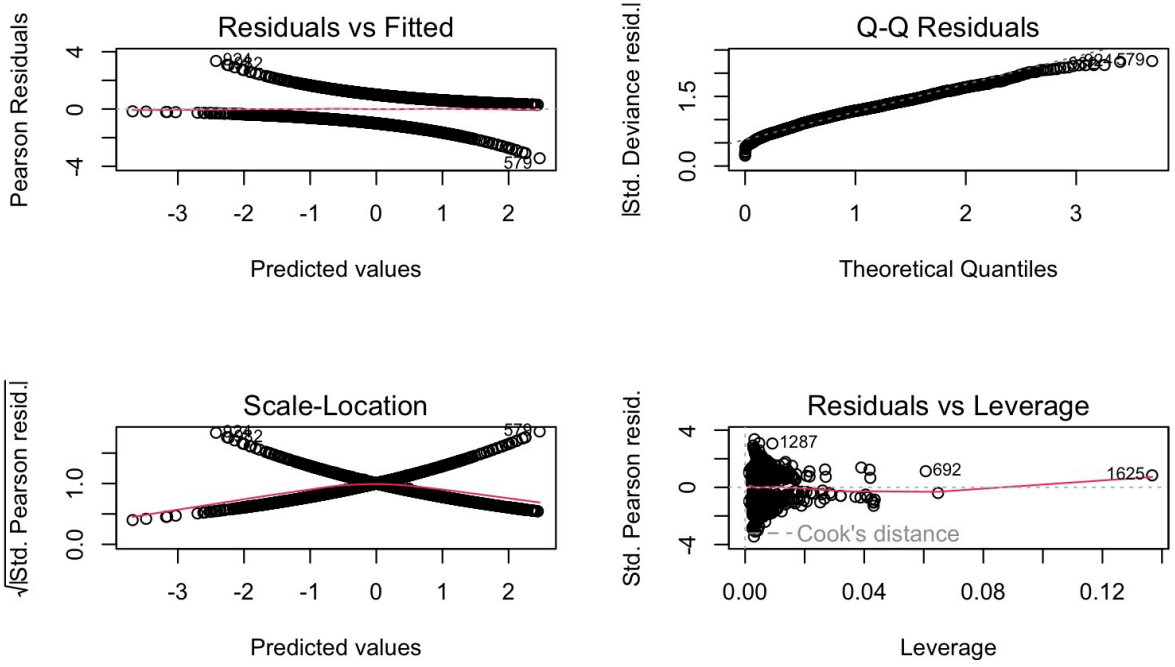
Replication

Original paper: only use **additive model**

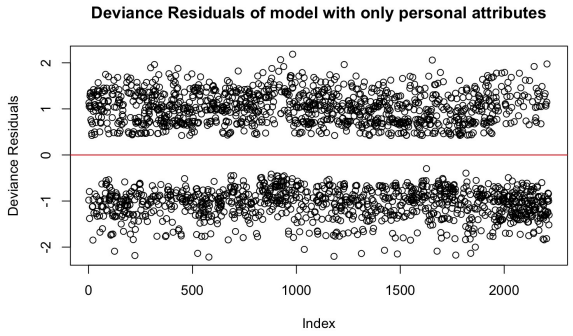
Replication study: consider an **interaction term** of **household wealth** and **owns home**

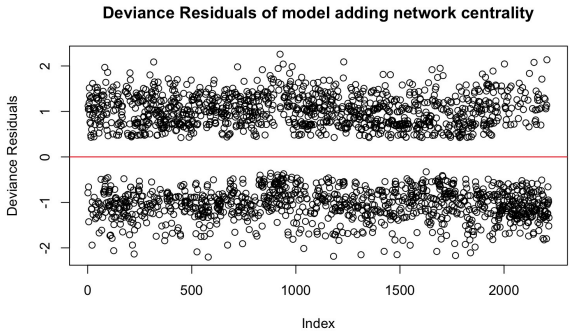
Replication-Check model assumptions

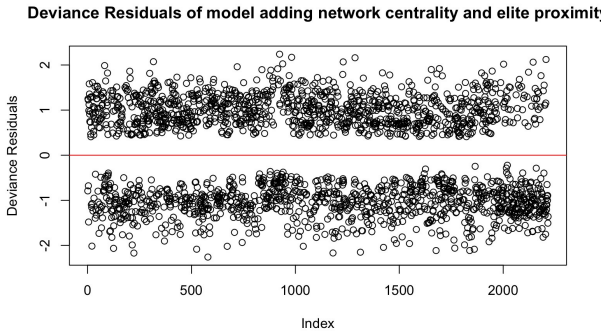
- First, let's check the assumptions:



Replication-Check model assumptions







Replication-Fit new models

```
# MODEL 4
alexandria_turnout_4 <- paste("turnout ~ ",
                    new_controls, " + hhwealthlog*owner ")  %>% formula()  %>%
 glm(data = alexandria_data,
    family = binomial(link = "logit"))

# MODEL 5
alexandria_turnout_5 <- paste("turnout ~ ",
                    new_controls,
                    "+ z1ev + hhwealthlog*owner")  %>% formula()  %>%
 glm(data = alexandria_data,
    family = binomial(link = "logit"))

# MODEL 6
alexandria_turnout_6 <- paste("turnout ~ ",
                    new_controls,
                    "+ z1ev + z1elite_avgprox + hhwealthlog*owner")  %>% formula()
%>% glm(data = alexandria_data,
    family = binomial(link = "logit"))
```

Replication
-Compare with original models

Table 1: **Estimates from a Model Predicting an Individual's Probability of Voting Using his SocialProximity to Elites and Other Variables (Based on the Statewide Elections in 1859 Alexandria)**

	Original			Interaction term		
	(1)	(2)	(3)	(1)	(2)	(3)
(Intercept)	-13.500*	-13.806*	-14.244*	-13.781*	-14.046*	-14.486*
	(2.146)	(2.148)	(2.167)	(2.157)	(2.157)	(2.178)
hhwealth	-0.011*	-0.010*	-0.010*	-0.015*	-0.014*	-0.014*
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
hhwealthlog	0.063*	0.041	0.043*	0.060*	0.039	0.041
	(0.021)	(0.022)	(0.022)	(0.021)	(0.022)	(0.022)
midstatus	0.687*	0.417*	0.290*	0.674*	0.416*	0.287*
	(0.103)	(0.119)	(0.123)	(0.103)	(0.120)	(0.124)
highstatus	0.864*	0.633*	0.587*	0.823*	0.603*	0.555*
	(0.186)	(0.193)	(0.194)	(0.188)	(0.195)	(0.196)
owner	0.765*	0.802*	0.781*	0.535*	0.583*	0.556*
	(0.150)	(0.151)	(0.152)	(0.164)	(0.166)	(0.167)
age	-0.091*	-0.097*	-0.099*	-0.096*	-0.102*	-0.104*
	(0.020)	(0.020)	(0.021)	(0.021)	(0.021)	(0.021)
agelog	4.378*	4.615*	4.738*	4.517*	4.730*	4.854*
	(0.808)	(0.810)	(0.817)	(0.813)	(0.815)	(0.822)
church	0.881*	0.879*	0.582*	0.851*	0.852*	0.551*
	(0.106)	(0.107)	(0.125)	(0.107)	(0.107)	(0.126)
usborn	0.723*	0.606*	0.543*	0.718*	0.607*	0.543*
	(0.119)	(0.123)	(0.124)	(0.119)	(0.123)	(0.124)
z1ev		-0.162*	-0.235*		-0.153*	-0.227*
		(0.038)	(0.041)		(0.038)	(0.041)
z1elite_avgprox			0.557*			0.562*
			(0.125)			(0.126)
hhwealthlog:owner				0.328*	0.301*	0.305*
				(0.101)	(0.102)	(0.102)
AIC	2697.508	2680.728	2659.522	2688.383	2673.522	2652.034
BIC	2754.543	2743.466	2727.964	2751.121	2741.963	2726.179
Log Likelihood	-1338.754	-1329.364	-1317.761	-1333.191	-1324.761	-1313.017
Deviance	2677.508	2658.728	2635.522	2666.383	2649.522	2626.034
Num. obs.	2216	2216	2216	2216	2216	2216

* $p < 0.05$

Conclusion:
For households that own property, holding other variables constant, for each unit increase in the logarithm of household wealth, the odds of an individual voting increase by approximately 35.66% compared to when the household does not own property. (The last Model)

Replication
-Compare with original models

```
# Perform an F-test to compare the two models
anova(alexandria_turnout_3, alexandria_turnout_6, test = "Chisq")
```

```
Analysis of Deviance Table

Model 1: turnout ~ hhwealth + hhwealthlog + midstatus + highstatus + owner +
    age + agelog + church + usborn + z1ev + z1elite_avgprox
Model 2: turnout ~ hhwealth + hhwealthlog + midstatus + highstatus + owner +
    age + agelog + church + usborn + z1ev + z1elite_avgprox +
    hhwealthlog * owner
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      2204     2635.5
2      2203     2626.0  1   9.4881 0.002068 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion: Adding an interaction term has better performance.