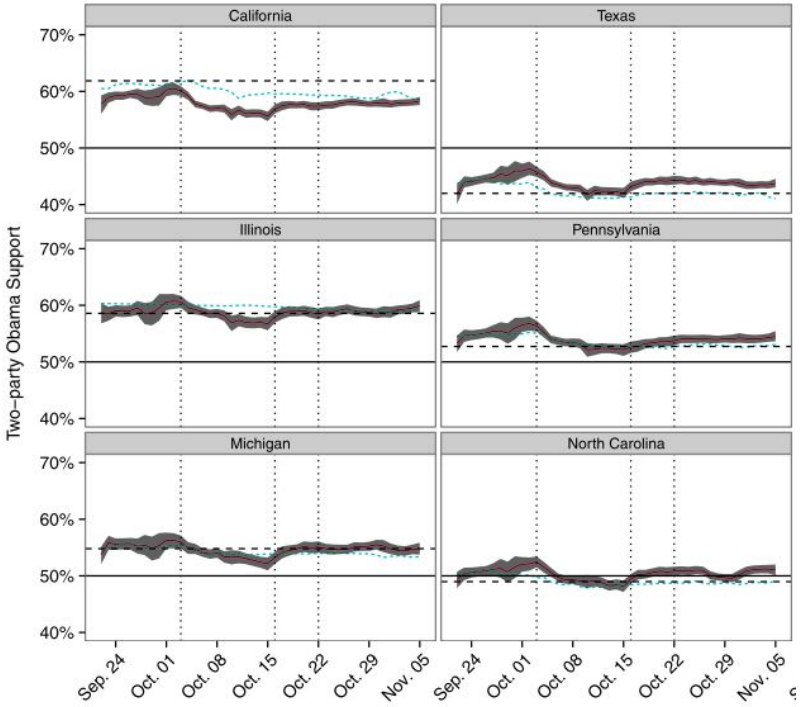
# Introduction to multilevel modelling



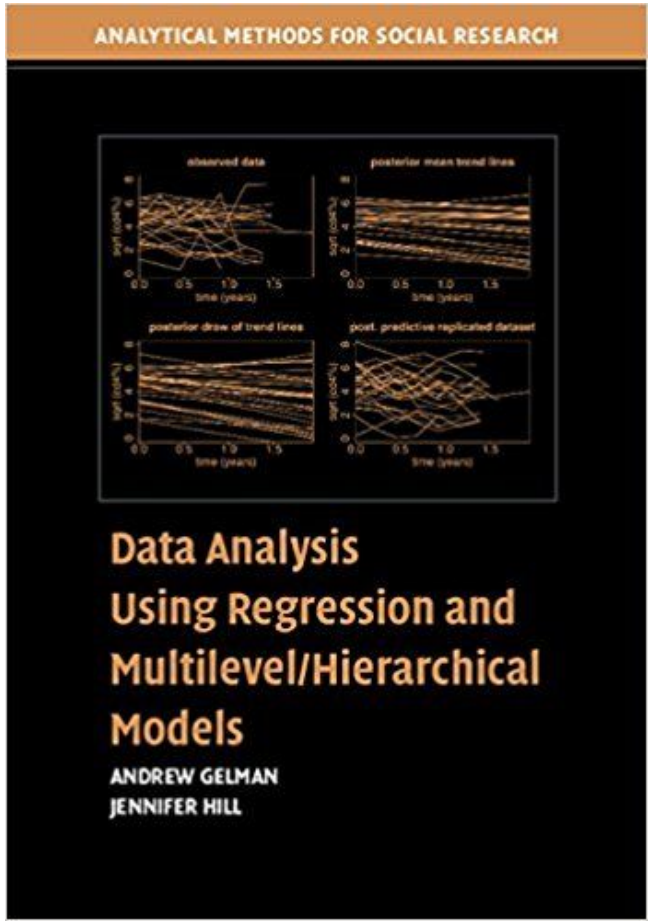

Wang, W., et al., Forecasting elections with non-representative polls. International Journal of Forecasting (2014)

# sources

**Gelman, Andrew, and Jennifer Hill. Data analysis using regression and multilevel/hierarchical models. Cambridge university press, 2006.**

https://stats.idre.ucla.edu/spss/seminars/spss-mixed-command/

https://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/reviewspss.pdf



# linear regression

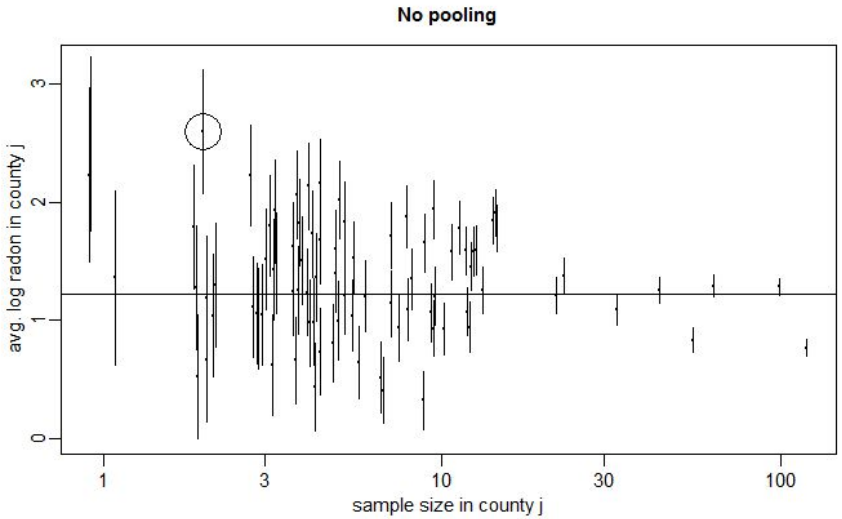
$$y = \alpha + \beta x + \text{error}$$

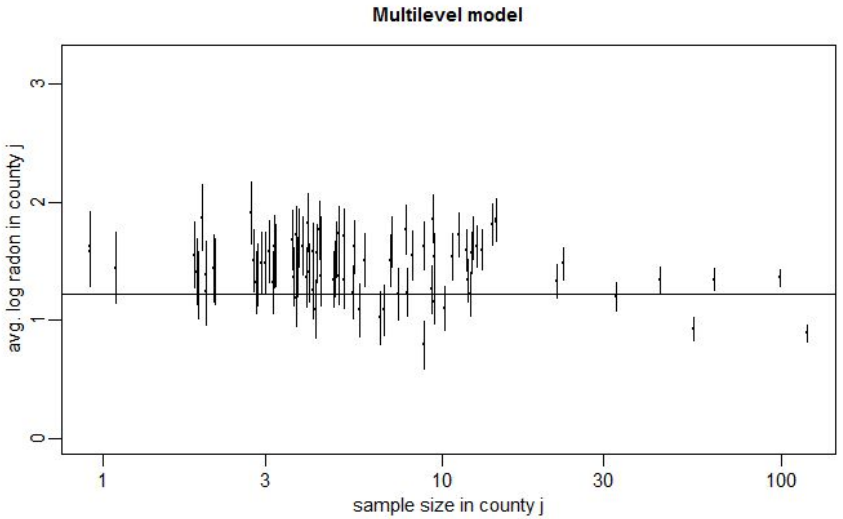
separate linear regressions for county j = 1,...,J

'no pooling' solution

$$y_j = \alpha_j + \beta_j x + \text{error}$$

# linear regression * J: no pooling



single model with county as categorical predictor

'complete pooling' solution

$$y = \alpha + \beta x_1 + \beta_{\text{country}_j} + \text{error}$$

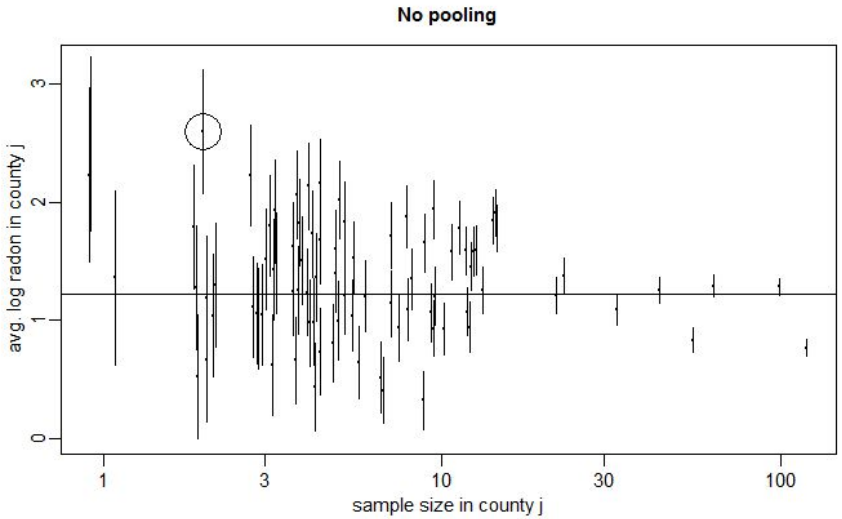
## random intercepts model: 'partial pooling'

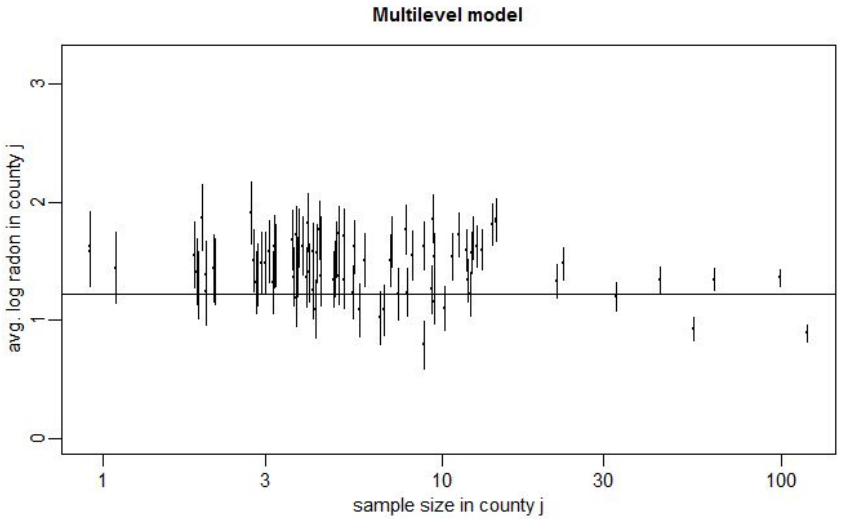
$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i,$$

for households $i = 1, ..., N$

for counties $j = 1, ..., J$

# random intercepts model: 'partial pooling'



# random intercepts model with group-level predictors

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i, \qquad \text{for households } \; i = 1, ..., N$$

$$\alpha_j = a + b u_j + \eta_j, \qquad \text{for counties } \; j = 1, ..., J$$

# random intercepts & random slopes

$$y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i,$$

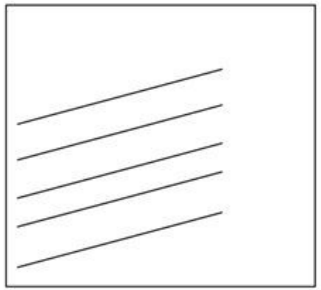
for households $i = 1, ..., N$

$$\alpha_j = a_0 + b_0 u_j + \eta_{j1},$$

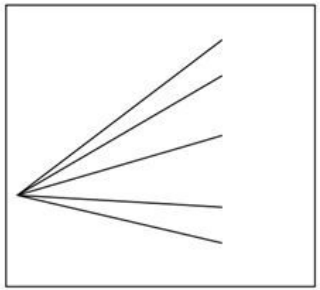
for counties $j = 1, ..., J$

$$\beta_j = a_1 + b_1 u_j + \eta_{j2},$$

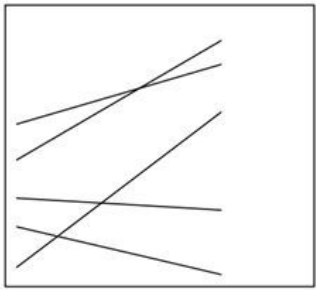
**Varying intercepts**



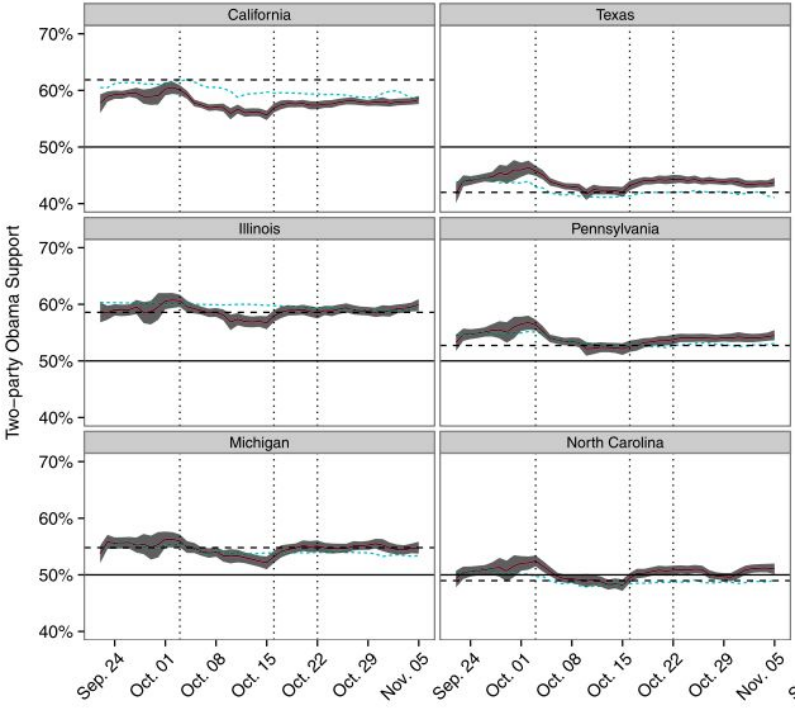
**Varying slopes**



**Varying intercepts and slopes**



# further examples from our research




Wang, W., et al., Forecasting elections with non-representative polls. International Journal of Forecasting (2014)

Figure 1: Sample imbalance

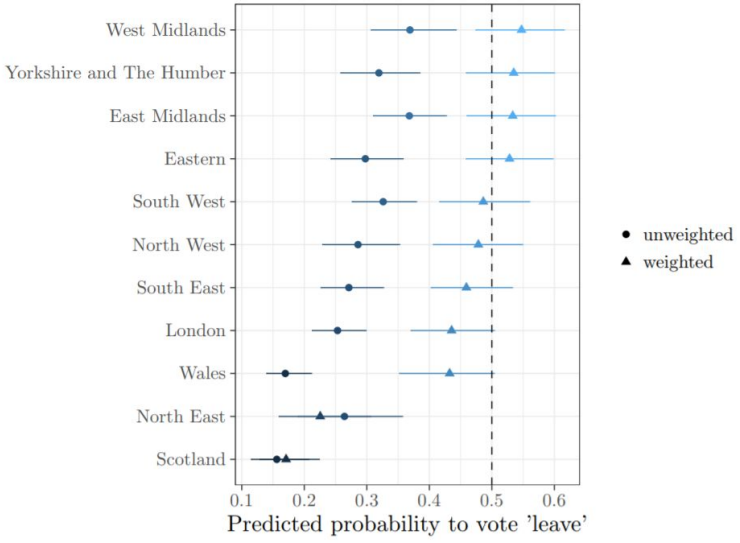
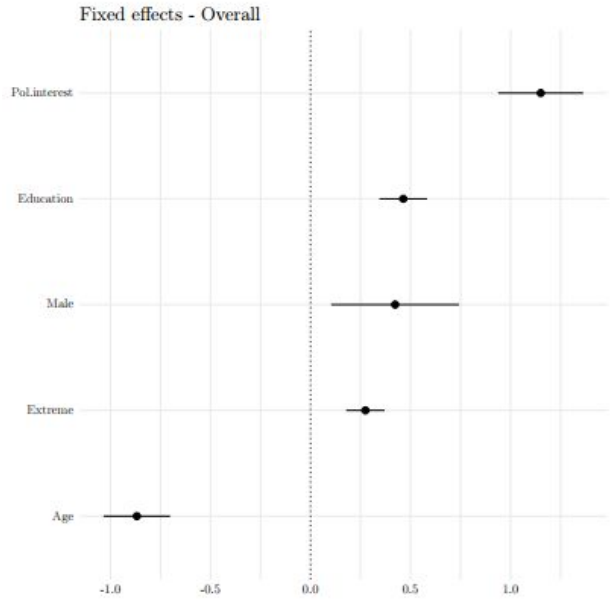



Figure 2: Random intercepts per region

Antonucci. L., et al., Challenging the narrative of the 'left behind' Brexiter. Competition and Change (2017)




| | Overall.ipc |
|---|---|
| **Individual level** | |
| Age | 0.194*** |
| | (0.018) |
| Male | 0.076** |
| | (0.037) |
| Education | 0.274*** |
| | (0.015) |
| Political interest | 0.132*** |
| | (0.023) |
| **Party level** | |
| Niche | 0.291 |
| | (1.347) |
| Governmental status | 0.366 |
| | (1.117) |
| Centre | 2.913** |
| | (1.411) |
| Right | 2.072** |
| | (0.890) |

# Varying intercept linear regression

source: Gelman & Hill, 2007

```
mixed
  y with x
 /fixed = x
 /print = solution testcov
 /random intercept | subject(group)
```

# Varying intrcpt w/ group-level predictr

source: Gelman & Hill, 2007

```
mixed
   y with x u
 /fixed = x u
 /print = solution testcov
 /random intercept | subject(group)
```

# Varying intrcpt w/ varying slope

source: Gelman & Hill, 2007

```
mixed
   y with x u
 /fixed = x u x*u
 /print = solution testcov
 /random intercept | subject(group) covtype(un)
```

# Reshape data set for 'stacking'

```
varstocases
 /id = id
 /make score from read write math science
 /index = subject(score)
 /keep = school
 /null = keep.
```