

# Predicting Constituency Vote Shares from Pre-Election Polls

Chris Hanretty (UEA)  
Benjamin E. Lauderdale (LSE)  
Nick Vivyan (Durham University)

# #1: The problem

# Constituency-level election prediction in the UK

- Generating constituency-level polling estimates for the 632 (England, Wales, Scotland) constituencies is infeasible.
  - For a sample of 500 per constituency, would need a national sample of 316,000.
- Uniform national swing is a reasonable approximation, but could be wrong in any given election.
- How can we combine national polling data and other sources of relevant information to generate better constituency-level predictions?

## #2: Using information about constituencies

# Constituency-level information about constituencies

- Principle:
  - People who live in constituencies with similar characteristics are more similar in their voting intentions
  - Each respondent we poll tells us a little bit about respondents in similar constituencies

# Constituency-level information about constituencies

- Principle:
  - People who live in constituencies with similar characteristics are more similar in their voting intentions
  - Each respondent we poll tells us a little bit about respondents in similar constituencies
- Procedure: Multilevel Regression
  - Build a regression model to predict individual-level votes with constituency-level characteristics
  - Use regression estimates to predict vote shares in each constituency

# Constituency-level information about constituencies

- Principle:
  - People who live in constituencies with similar characteristics are more similar in their voting intentions
  - Each respondent we poll tells us a little bit about respondents in similar constituencies
- Procedure: Multilevel Regression
  - Build a regression model to predict individual-level votes with constituency-level characteristics
  - Use regression estimates to predict vote shares in each constituency
- Caveats:
  - Only as helpful as the predictive power of the variables we use
  - Vote in last election is very powerful (near uniform swing)

# Individual-level information about constituencies

- Principle:
  - People who share demographic characteristics are more similar in their voting intentions
  - Each respondent we poll tells us a little bit about respondents with similar characteristics



# Individual-level information about constituencies

- Principle:
  - People who share demographic characteristics are more similar in their voting intentions
  - Each respondent we poll tells us a little bit about respondents with similar characteristics
- Procedure: Multilevel Regression + Post-stratification (MRP)
  - Build a regression model to predict individual-level votes with individual-level characteristics
  - Use Census data to determine how many of each type of person is in each constituency (construct post-stratification weights)
  - Use regression estimates plus post-stratification weights to predict vote shares in each constituency

# Individual-level information about constituencies

- Principle:
  - People who share demographic characteristics are more similar in their voting intentions
  - Each respondent we poll tells us a little bit about respondents with similar characteristics
- Procedure: Multilevel Regression + Post-stratification (MRP)
  - Build a regression model to predict individual-level votes with individual-level characteristics
  - Use Census data to determine how many of each type of person is in each constituency (construct post-stratification weights)
  - Use regression estimates plus post-stratification weights to predict vote shares in each constituency
- Caveats:
  - Only as helpful as the predictive power of the variables we use
  - UK Census data availability/categories are a constraint

# Geographic information about constituencies

- Principle:
  - People in nearby constituencies are more similar in their voting intentions
  - Each respondent we poll tells us a little bit about respondents in nearby constituencies

# Geographic information about constituencies

- Principle:
  - People in nearby constituencies are more similar in their voting intentions
  - Each respondent we poll tells us a little bit about respondents in nearby constituencies
- Procedure: Spatially Correlated Random Effects (SCRE)
  - Build a regression model where the constituency-level random effects are spatially correlated
  - Use regression estimates to predict vote shares in each constituency

# Geographic information about constituencies

- Principle:
  - People in nearby constituencies are more similar in their voting intentions
  - Each respondent we poll tells us a little bit about respondents in nearby constituencies
- Procedure: Spatially Correlated Random Effects (SCRE)
  - Build a regression model where the constituency-level random effects are spatially correlated
  - Use regression estimates to predict vote shares in each constituency
- Caveats:
  - Only as helpful as the predictive power of geography

# More information is better

- We don't need to choose between individual, constituency, and geographic data
- We can combine all three.

# #3: Revisiting 2010

# Survey data

- 2010 British Election Study CIPS data (un-weighted)
- 12,177 total sample size
- 632 constituencies in England, Wales and Scotland
- 19.3 mean respondents per constituency is (range: 3 to 46)
- How well could we have predicted the 2010 constituency-level results given these data?



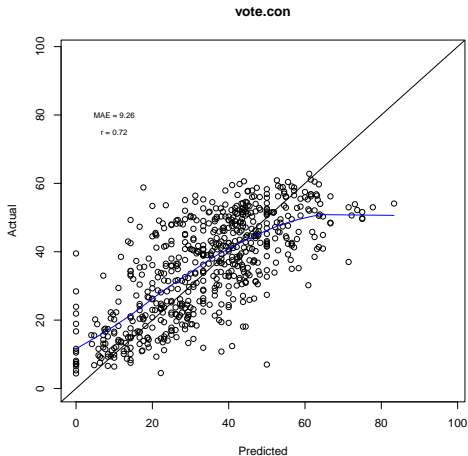
## Actual and survey-based national vote shares

Party	Actual vote share	Raw survey vote share
Conservatives	36.1	35.6
Labour	29.0	26.0
Liberal Democrats	23.0	27.1

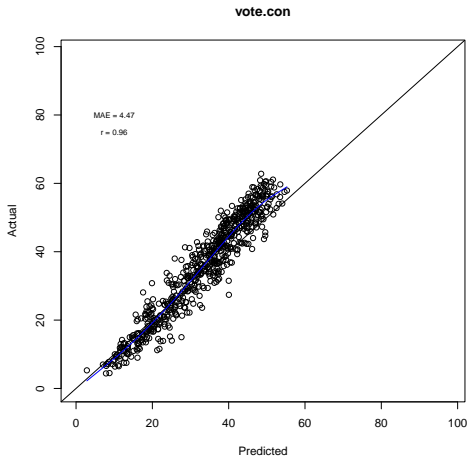
# Individual, constituency, and geographic data

- Constituency-level (UK Census)
  - Lagged vote shares (2005 on 2010 boundaries)
  - Log population density
  - Log of median earnings
  - Religious composition (4 levels)
  - Region (11 levels)
- Individual-level (UK Census)
  - Male/Female
  - Renter/Owner
  - Private/Public Sector
  - Age Group (8 levels)
  - Education Qualifications (6 levels)
  - Social Grade (4 levels)
- Geographic (UK Ordnance Survey)
  - Constituency adjacency

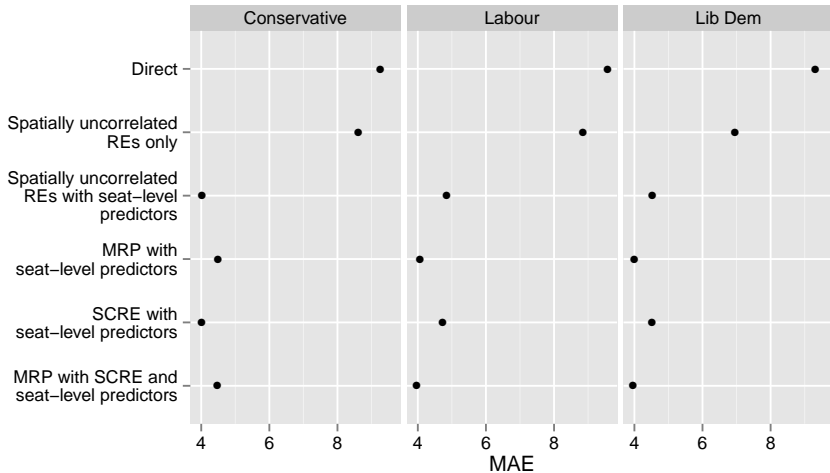
# Predicted vs Actual Conservative Vote Estimated using disaggregation



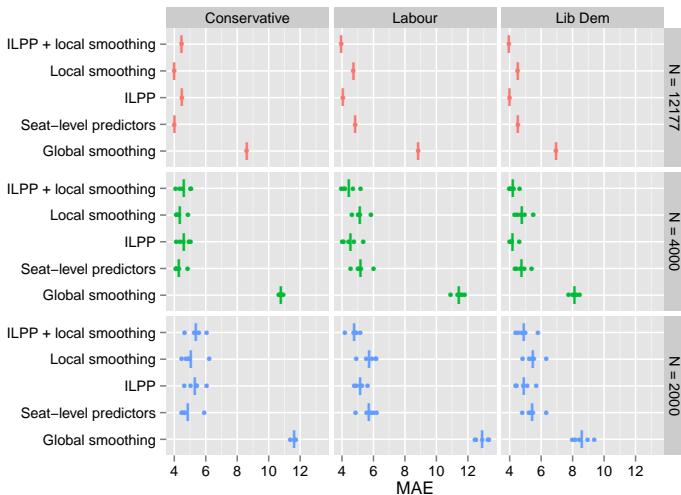
# Predicted vs Actual Conservative Vote Estimated using MRP with SCRE



# Model comparison



# Robustness to reduced sample size



## #4: Conclusion

# Summary

- Multiple kinds of information can be used to generate constituency-level estimates of vote intention from a medium-sized national poll.
  - Individual-level demographics
  - Constituency-level characteristics
  - Geographic proximity
- The payoff is large: 10x-100x improvement in the effective sample size.
- <http://constituencyopinion.org.uk/>