Compositional data analysis of vote shares in the 2001 Australian Federal election

Derek Chong, Sinclair Davidson
Lisa Farrell and Tim R.L. Fry

Context

- Geospatial modelling has investigated spatial voting patterns in Federal elections
- We are interested in investigating the determinants of voting outcomes at the electoral division level
- We hypothesise that voting outcomes are related to the characteristics of the electoral divisions
- We will focus upon first preference voting for The House of Representatives in the 2001 Federal election

Data

- 2001 Federal election took place three months after the 2001 census
- Thus we have an ideal match between the voting data and census data on the characteristics of the electoral divisions.
- We use Australian Electoral Commission data for the 150 electoral divisions on the vote shares of six parties
  - Coalition, ALP, Democrats, Greens, ONP and Others
- This is matched to census data on characteristics of the electoral division

Compositional data analysis

- Use the additive log-ratio transformation and fit a multivariate regression model to the transformed data:

$$y_i = \ln\left(\frac{s_i}{s_n}\right) = \mu_i\left(\mathbf{Z}, \theta\right) + u_i = Z_i' \theta_i + u_i,$$

$$\mathbf{u} = \left[u_i\right] \sim N\left(\mathbf{0}, \Sigma\right).$$

- Can invert to return to vote shares.
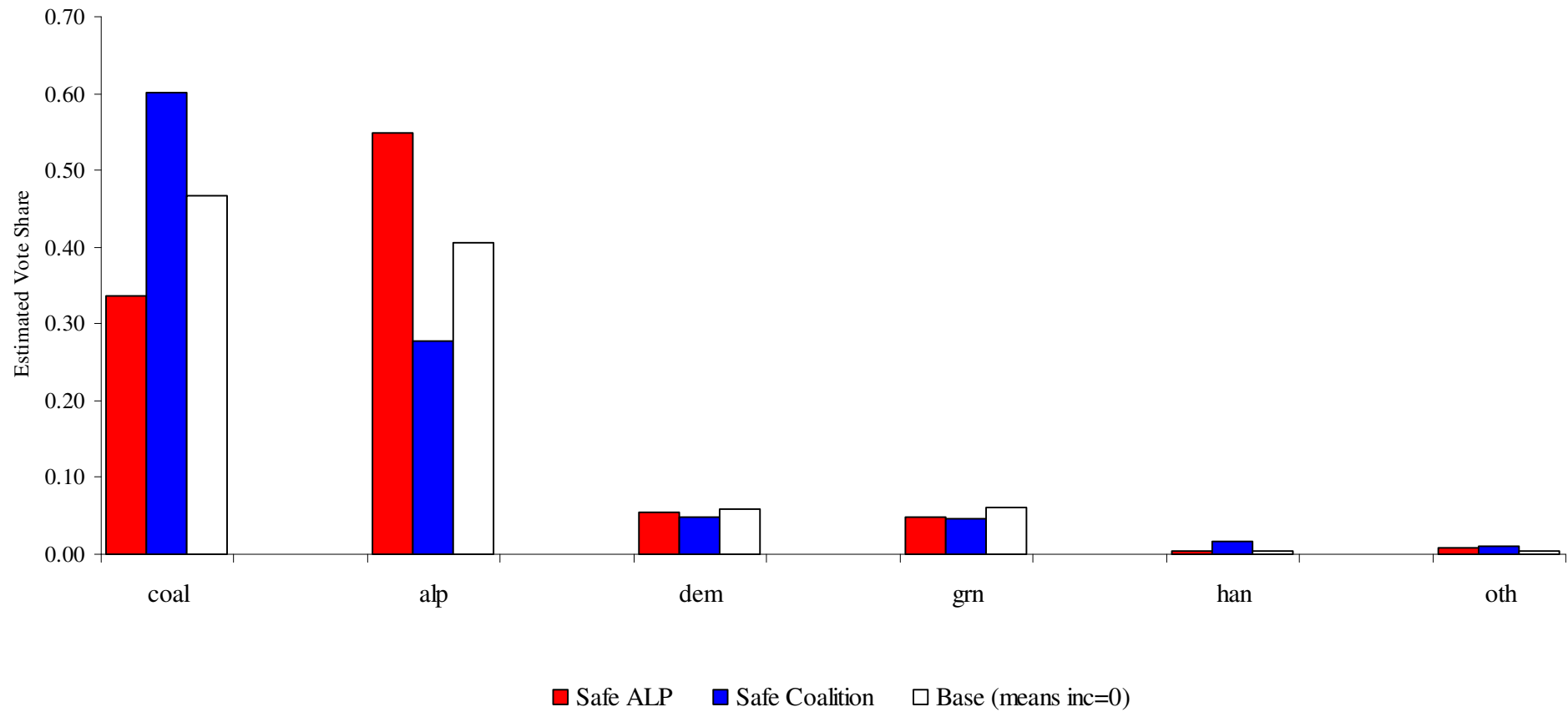
Zeros in compositional data analysis

- Cannot take the log of zero
- Solution is to use "modified Aitchison" with minimum (sensible) replacement values
- We will assume that when we observe a zero we can replace it with a one
  - That is, should the parties have stood then the candidates would have voted for themselves
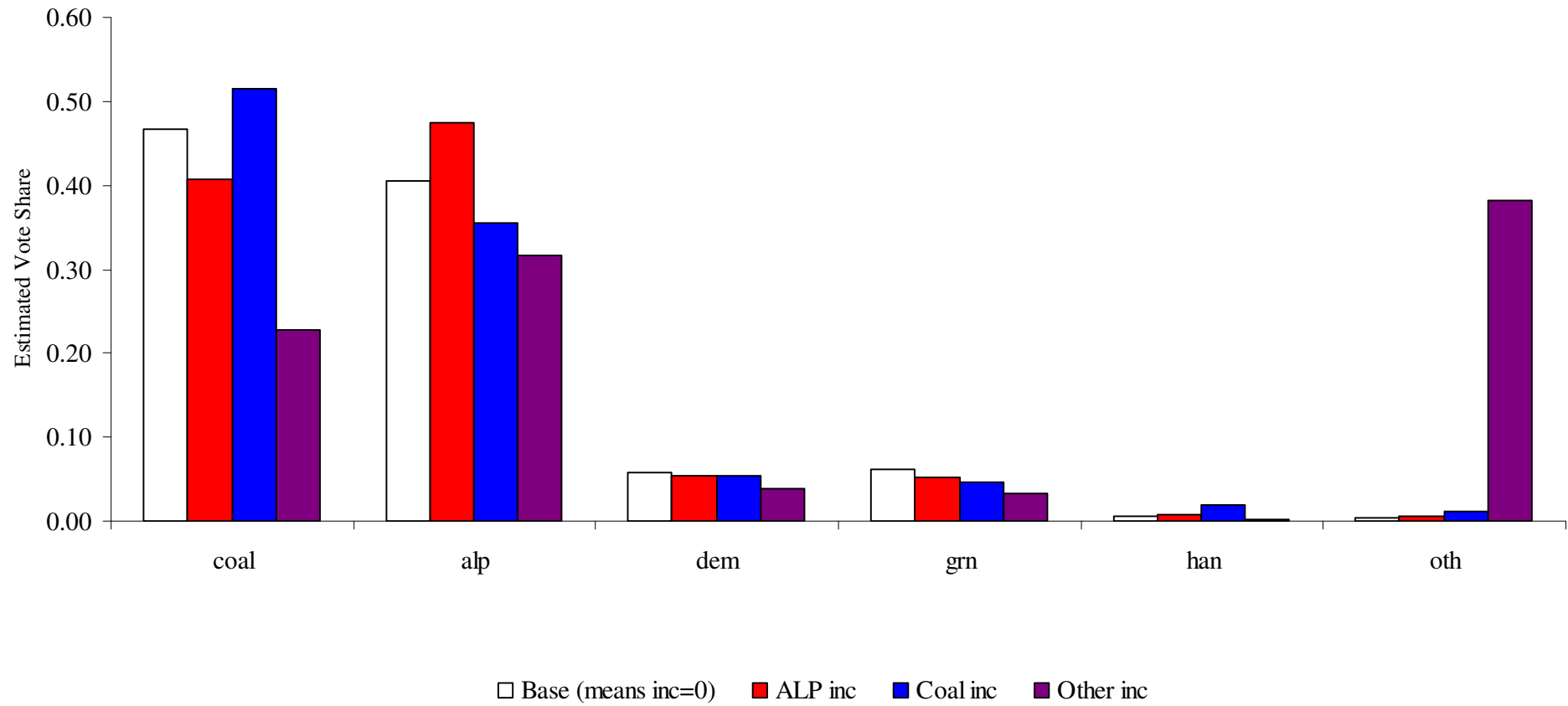- Check for robustness of estimates to the zero replacement procedure

Results

- Estimates make sense and are well determined.
- Robust to changing the the zero replacement (1, 2, 3, 5, 10, 20 or 50 votes)
- Correctly classify 139 of 150 electoral divisions
  - New England predicted Coalition actually Independent
    - AEC classified as Safe National Party
  - Of the six predicted ALP but actually Coalition <u>all</u> were classified by the AEC as marginal ALP
  - Of the four predicted Coalition but actually ALP three were classified by the AEC as marginal (ALP and Liberal)

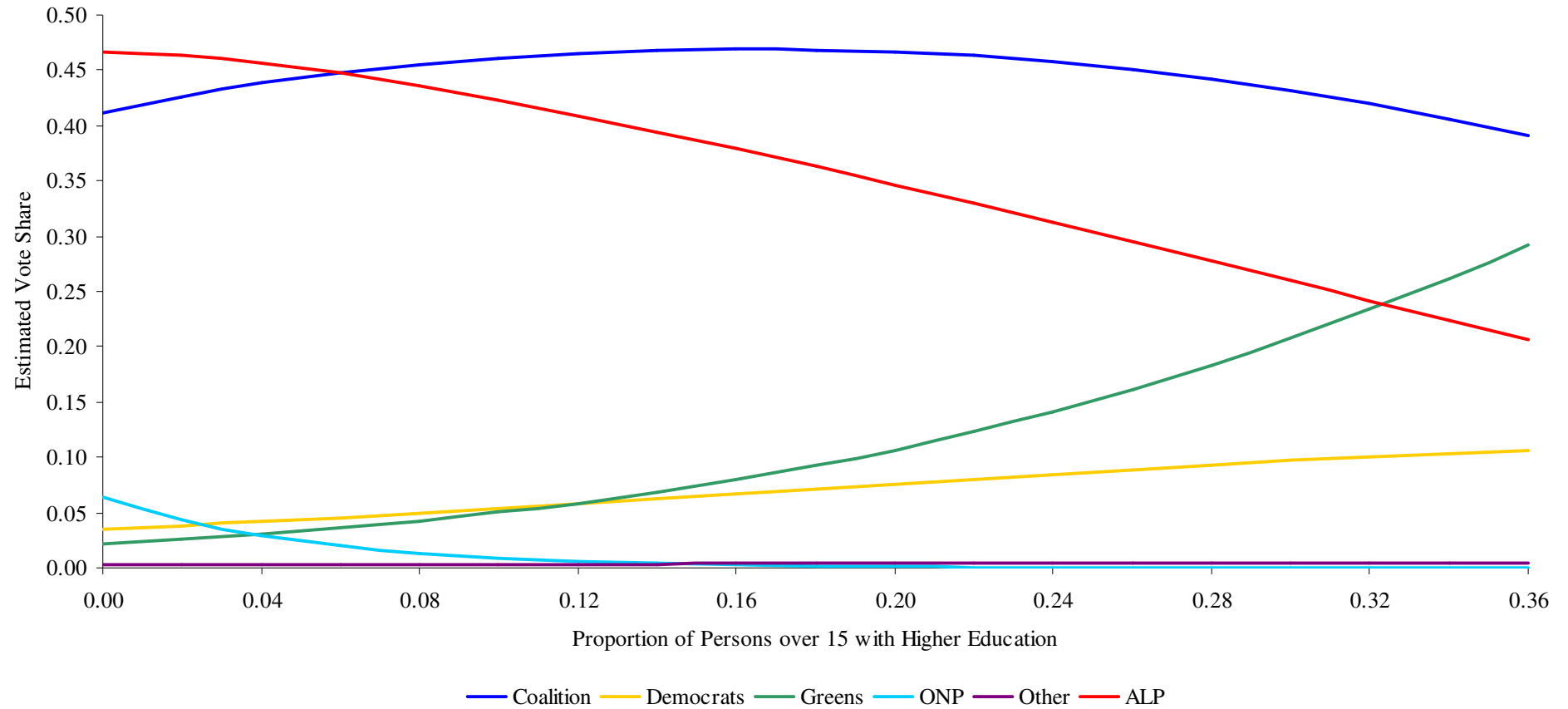# Estimated Vote Shares for Stylised Electoral Divisions

Chart showing Estimated Vote Share (y-axis, 0.00 to 0.70) for categories: coal, alp, dem, grn, han, oth. Series: Safe ALP (red), Safe Coalition (blue), Base (means inc=0) (white).
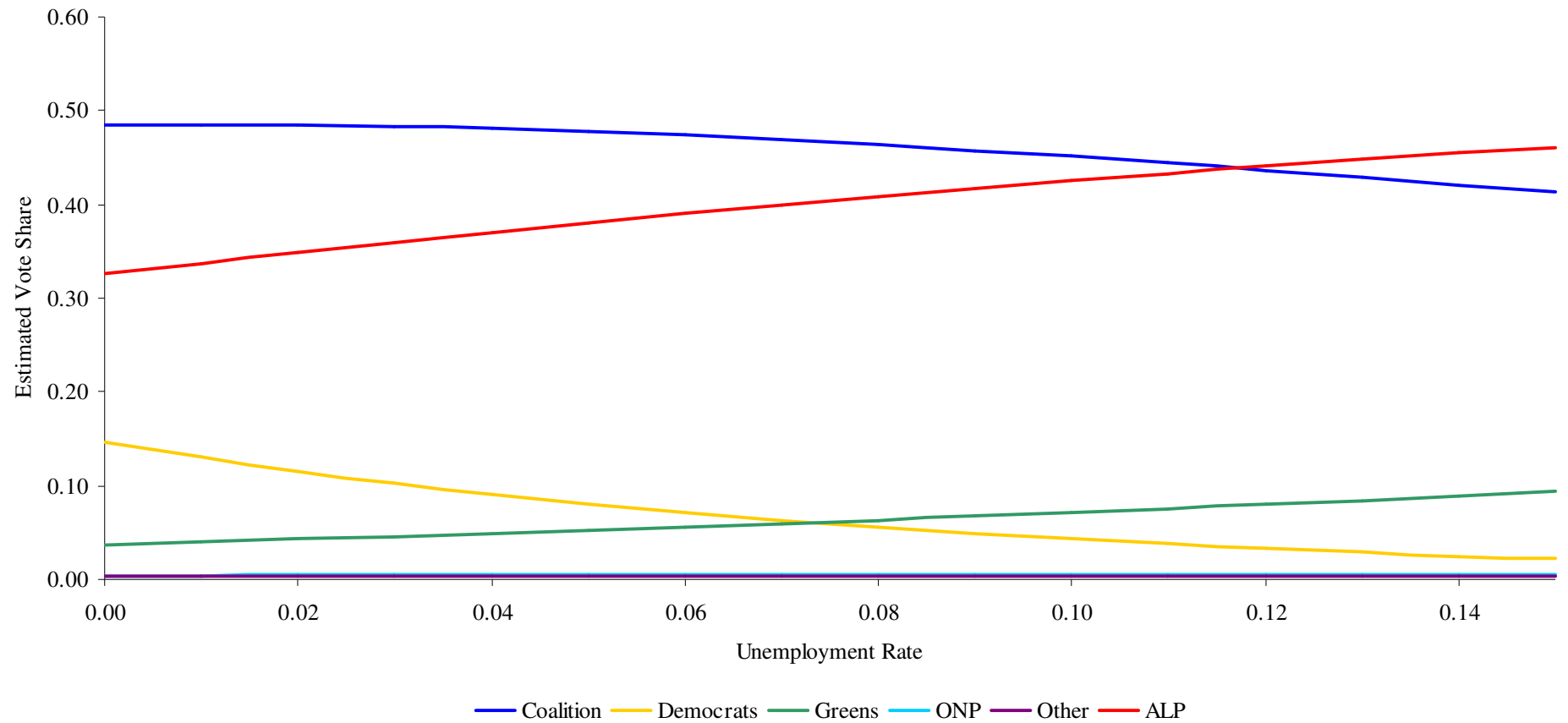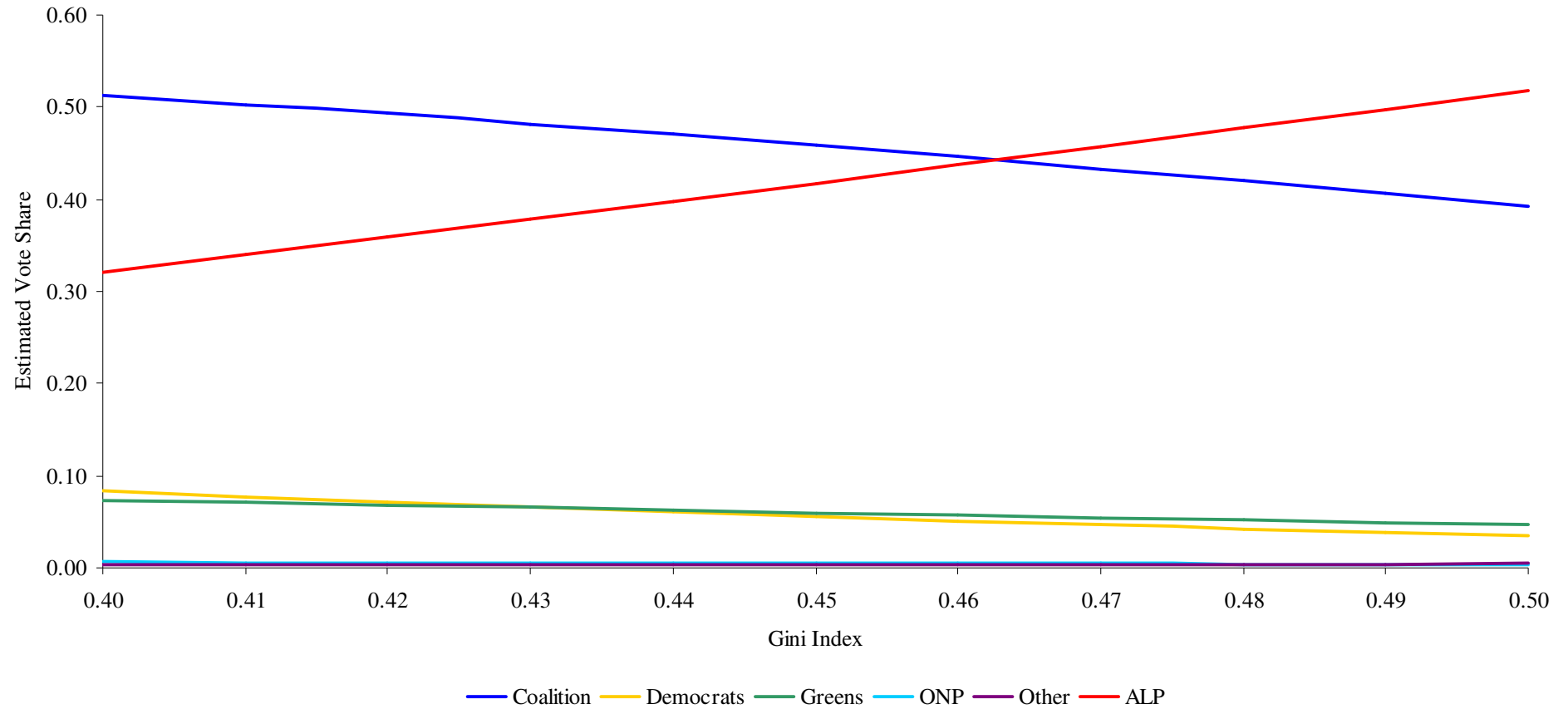
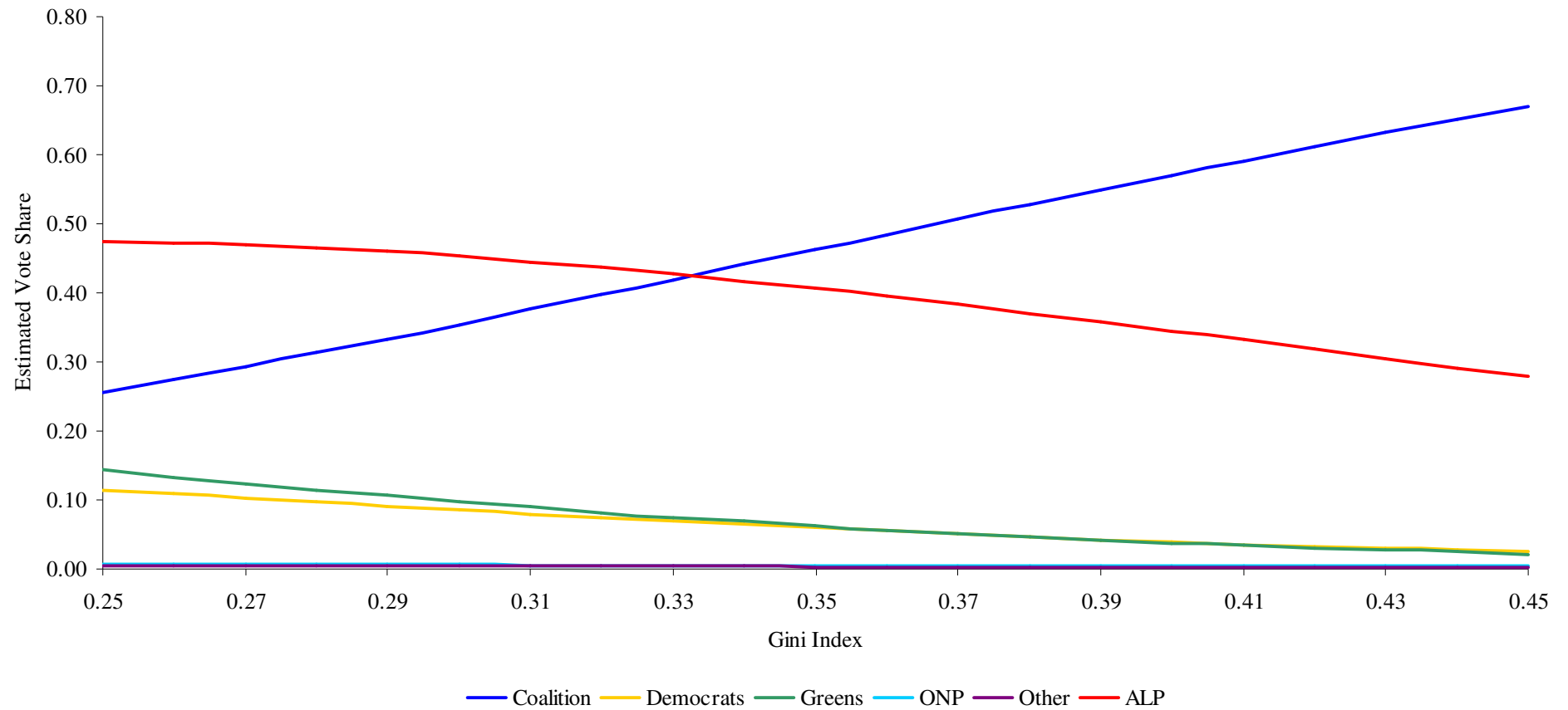Estimated Incumbency Effect

Estimated Impact of Higher Education

Estimated Impact of Unemployment

Estimated Impact of Income Inequality

Estimated Impact of Age Inequality

Results

- We can use our estimated model to give us the estimated votes by electoral division
  - For a party that did not contest an electorate (e.g. ONP) this is an estimate of the vote that they might have obtained had they stood
  - IF the party (ONP) makes rational decisions then the vote in electorates that it does contest is larger than that in which it does not stand a candidate.
    - t-test statistic is 8.8179

Conclusions

- First application of compositional data analysis techniques to Australian election data

- Still need to fine tune the work – but results are encouraging

- Also working on a number of related issues concerning functional forms and zero replacement in compositional data analysis