Analytical Strategies for Comparative Electoral Research: Stacks and Y-hats

Cees van der Eijk

1st European Conference on Comparative Electoral Research
University of National and World Economy, Sofia, December 1-3, 2011

The problems

The University of Nottingham

- Dependent variable is different
  - between elections in a single country
  - between countries / political systems
- Dependent variable is nominal-level
- Independents are measured differently
  - wording differences
  - response format / # of response options

Different Dependent Variables

The University of Nottingham

- Traditional ‘solutions’
  - Recoding of parties into left/right positions
  - Recoding into government/opposition
  - Recoding into, e.g., ‘socialist/else’, ‘liberal/else’, etc.
- Disadvantages:
  - Degradation of empirical data (loss of information)
  - Classification problems (e.g., is a party ‘left’ or not?)
  - Implausible assumption: other aspects of parties are irrelevant for choice (→ improperly specified models)

Analysing the Dependent Variable
(traditional)

The University of Nottingham

- Ubiquitous : multinomial logistic regression for each dataset separately. Disadvantages:
  - Epistemological shortcomings:
    only voter characteristics can be used as IVs, party characteristics (e.g., party size) or voter-party relationships (e.g., ideological affinity) cannot be included as independent variables
  - With increasing # of parties: increasingly less intelligible (e.g., an 8 party system yields 28 binary comparisons, each with their own coefficients)
  - Does not allow pooled analyses for different contexts

Analysing the Dependent Variable
(innovative)

The University of Nottingham

- Preferred method: conditional logistic regression
  - Replaces individuals by individual*party as cases (→ stacked data), DV is dichotomous (0/1) and ipsative
  - Allows use of party characteristics and voter-party relationships as independent variables
  - Stacked data for different elections / different countries can be pooled ('add cases'), allowing for explicit modelling of context characteristics
  - Analytical design similar as for conditional OLS for non-ipsative preference ratings (see, e.g., van der Eijk et al 2006)

The IV problem

The University of Nottingham

- Measurement differences
  - Question wording / # of questions
  - Wording and number of response options
- Traditional ‘solution’: recoding to lowest common denominator. Disadvantages:
  - Degradation of empirical data (loss of information)
  - Classification problems

The Independent Variable problem – Y-hat approach

Replace the IVs by new measures, expressed on the same scale as the stacked DV by an appropriate regression model for each of the separate parties (Y):

$$Y = a + bX + e$$

$$\text{Y-hat} = a + bX$$

Then stack the Y-hats to yield IV in stacked dataset

NB: correlation between Y and X is identical to correlation between Y and Y-hat: hence no information is 'lost', nor is anything added that changes the original relationships between dependent van independent variables

The Independent Variable problem – Y-hat approach - 2

The University of Nottingham

Advantages of Y-hat approach

- No degradation of empirical information (the transformation is loss-free)
- Fits with stacked and pooled data structures
- Can be combined with extra variables with additional information about the original X variables (allowing simple and straightforward assessment of, e.g., differences in response format)

Analyses can vary:

- Party level centering before stacking of Y-hats, or alternatively specification of party level fixed or random effects in explanatory model, etc.

References

Eijk, Cees van der, and Mark N. Franklin, 2006. *Choosing Europe?* Ann Arbor: University of Michigan Press
Chapter 20 of this book contains an extensive discussion of the disadvantages of traditional ways of analysing party choice in different contexts, and discusses in an intuitive way the logic and advantages of the Y-hat approach. Buried in many footnotes is additional information of a validating nature.

Eijk, C. van der, W. van der Brug, M. Kroh & M.N. Franklin, 2006. "Rethinking the Dependent Variable in Voting Behavior – On the Measurement and Analysis of Electoral Utilities", *Electoral Studies*, 25, 424-47.
This article discusses the stacking and Y-hat approach as alternative to discrete choice models, discusses (partly in footnotes) some not well-known disadvantages of discrete choice models, and provides a large number of references to a variety of substantively different applications of the 'Stack&Y-hat' approach.
In addition to the 'Stack&Y-hat' approach, the article also discusses at length the PTV approach as a basis for studying the calculus of voting.

Brug, W. van der, C. van der Eijk and M. Franklin, 2007. *The Economy and the Vote*. Cambridge: Cambridge University Press
This book contains the most eleborate application of the 'Stack&Y-hat' approach, and addresses economic voting questions. It combines comparisons over time and comparisons across countries in a single model, and exploits the 'Stack&Y-hat' approach for the construction of relevant counterfactual models.