Analyzing electoral utilities (PTV’s) in the EES’s

Cees van der Eijk
cees.vandereijk@nottingham.ac.uk
Mark Franklin
mark.franklin@eui.eu
We have a number of parties in <country> each of which would like to get your vote. How probable is it that you will ever vote for the following parties? Please specify your views on a 10-point-scale where 1 means "not at all probable" and 10 means "very probable". You may use any number between 1 and 10 to specify your views. If you think of Party 1: what mark out of ten best describes how probable it is that you will ever vote for Party 1? And what mark out of ten best describes how probable it is that you will ever vote for Party 2? And for Party 3, etc.

These are known as the Propensity To Vote (PTV) questions.
Some uses of PTV’s in electoral studies
(more detailed discussion in van der Eijk et al, 2006)

- Description of:
  - support base of political parties, particularly useful when studying small political parties
  - party competition: the extent to which the basis of electoral support for different parties overlaps
  - voter variables: the extent to which an individual voter is ‘torn’ between different parties, subject to competition

These descriptions can be obtained by appropriate data-manipulation commands in statpacks, or by are provided by a STATA routine called `utilstat` (available on EES website)

- As dependent variable replacing party choice in comparative analyses, thus overcoming problems of incomparability between party systems.
Support base of political parties

- Aggregating over respondents provides an estimate of the support base for a party
  - This requires an aggregation rule in the form of summation after recoding of PTV-scores, e.g.:
    - (1-7=0) (8-10=1)
    - (1-4=0) (5-6=0.5) (7-10=1)
    - (1=0) (2=0.11) (3=0.22) …..(8=0.78) (9=0.89) (10=1)
    - Etc

- To the extent that respondents give high PTV scores to several parties, the potential support for those parties will overlap – see next 2 slides for example based on Britain 1989
Overlapping support for parties in Britain, 1989
Voter variables based on PTV’s

- Degree to which a voter supports more than one party / is subject to party competition
- Can form basis for description of systems via aggregation (see example on next slide, from Kroh et al., 2007)
Voter variables based on PTV’s: being subject (or not) to competition

Table 11.2 Proportions of voters subject to intense competition and beyond competition (1989, 1994, and 1999)

<table>
<thead>
<tr>
<th>Country</th>
<th>1989 Subject to intense competition</th>
<th>1994 Subject to intense competition</th>
<th>1999 Subject to intense competition</th>
<th>1989 Beyond competition</th>
<th>1994 Beyond competition</th>
<th>1999 Beyond competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>-</td>
<td>-</td>
<td>0.37</td>
<td>-</td>
<td>-</td>
<td>0.36</td>
</tr>
<tr>
<td>Belgium-Flanders</td>
<td>0.41</td>
<td>0.39</td>
<td>0.33</td>
<td>0.39</td>
<td>0.36</td>
<td>0.34</td>
</tr>
<tr>
<td>Belgium-Walloon</td>
<td>0.26</td>
<td>0.30</td>
<td>0.38</td>
<td>0.52</td>
<td>0.45</td>
<td>0.30</td>
</tr>
<tr>
<td>Britain</td>
<td>0.29</td>
<td>0.28</td>
<td>0.42</td>
<td>0.45</td>
<td>0.45</td>
<td>0.28</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.37</td>
<td>0.39</td>
<td>0.40</td>
<td>0.34</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>Finland</td>
<td>-</td>
<td>-</td>
<td>0.56</td>
<td>-</td>
<td>-</td>
<td>0.22</td>
</tr>
<tr>
<td>France</td>
<td>0.58</td>
<td>0.53</td>
<td>0.62</td>
<td>0.19</td>
<td>0.20</td>
<td>0.14</td>
</tr>
<tr>
<td>Germany</td>
<td>0.29</td>
<td>0.33</td>
<td>0.34</td>
<td>0.39</td>
<td>0.34</td>
<td>0.36</td>
</tr>
<tr>
<td>Greece</td>
<td>0.23</td>
<td>0.30</td>
<td>0.42</td>
<td>0.63</td>
<td>0.49</td>
<td>0.36</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.50</td>
<td>0.47</td>
<td>0.52</td>
<td>0.26</td>
<td>0.28</td>
<td>0.17</td>
</tr>
<tr>
<td>Italy</td>
<td>0.36</td>
<td>0.45</td>
<td>0.53</td>
<td>0.42</td>
<td>0.27</td>
<td>0.23</td>
</tr>
<tr>
<td>Luxemburg</td>
<td>0.38</td>
<td>0.37</td>
<td>0.36</td>
<td>0.41</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.35</td>
<td>0.44</td>
<td>0.51</td>
<td>0.40</td>
<td>0.25</td>
<td>0.18</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.32</td>
<td>0.29</td>
<td>0.31</td>
<td>0.45</td>
<td>0.47</td>
<td>0.57</td>
</tr>
<tr>
<td>Spain</td>
<td>0.33</td>
<td>0.29</td>
<td>0.29</td>
<td>0.48</td>
<td>0.50</td>
<td>0.48</td>
</tr>
<tr>
<td>Sweden</td>
<td>-</td>
<td>-</td>
<td>0.42</td>
<td>-</td>
<td>-</td>
<td>0.26</td>
</tr>
<tr>
<td>Mean EU-12</td>
<td>0.36</td>
<td>0.37</td>
<td>0.42</td>
<td>0.41</td>
<td>0.35</td>
<td>0.31</td>
</tr>
<tr>
<td>Mean EU-15</td>
<td>-</td>
<td>-</td>
<td>0.43</td>
<td>-</td>
<td>-</td>
<td>0.30</td>
</tr>
</tbody>
</table>

- Subject to intense competition: difference between the two highest ranked parties is 0 or 1.
- Beyond competition: difference between the two highest ranked parties is more than 3.
Using PTV’s as dependent variable

- The philosophy of analyzing party choice via electoral utilities has been described in detail elsewhere:
  - Tillie 1995
  - van der Eijk & Franklin 1996 (Ch. 20)
  - van der Eijk et al. 2006

- The procedure described in this presentation is based on the use of SPSS.
  - In STATA you would use the Reshape (wide to long) command
Constructing a ‘stacked’ datafile

- If the data pertain to various countries (as is the case in EES data) the following procedure has to be performed for each country separately. If one would like to analyze the data from all countries simultaneously, this can be done by subsequently pooling the stacked data-files of the various countries (in SPSS by `merge files>add cases`).

- The procedure has to be performed simultaneously for all dependent and independent variables. If one wants to add another independent in a later stage, the process has to be started all over again, or variables can be added later using table lookup with respondent and party id as indicator variables.
Sequence of steps

1. Identify dependent and independent variables. The dependent variable, if it is a ‘generic ptv’ shouldn’t require any special treatment before stacking.

2. Insert in the unstructured dataset a set of variables for the identification of the stacks (i.e., in our case: parties).

3. If necessary: transform the independent variables into an appropriate form (how to is discussed later).

4. Use the **Restructure** option in the SPSS data menu for the actual stacking (**reshape** in STATA).
Identification of stacks

- Create as many identifying variables as there are parties. These variables have the same value for each respondent in the unstructured data-file (they are thus constants). In the case of, e.g., 4 parties:
  
  - compute p1=1.
  - compute p2=2.
  - compute p3=3.
  - compute p4=4.

- These variables will also be stacked, in order to yield a single identifier for parties in the stacked file. In the restructuring procedure in SPSS this stacked variable can be named at will.
Independent variables

- Determine for each independent variable of interest of what type it is:
  - Describing respondent characteristics
    - Sex, age, political interest, etc
  - Describing party characteristics
    - Size, government status, etc
  - Describing respondent-party relationship
    - Left-right distance, Respondent-party similarity, etc
Defining (respondent × party) variables

- **Distances**
  - i.e. between voter and each of the parties on the L/R scale, pro/anti EU scale (NB: define distances by *absolute* differences!)

- Theoretically constructed similarities, based on theoretical terms and contextual knowledge. For example, if religion is an important cleavage:
  - voter is religious AND party is religious: similarity=1
  - voter is not religious AND party is not religious: similarity=1
  - voter is not religious AND party is religious: similarity=0
  - voter is religious AND party is not religious: similarity=0

- Inductively generated (respondent × party) variables :
  - Y-hat procedure (see next sheet)
Perform the following operations in the unstructured data-matrix for each of the parties in turn:

- Regress electoral utility on the independent variable to be transformed
- Save the predicted value (the y-hat)
- Determine the mean of the y-hat in question
- Center the y-hat around 0 by subtracting mean
- Save, and use as one the variables to be stacked

This should for each independent variable yield as many centered y-hat variables as there are parties to be stacked, stacking implies that these are restructured into a single variable in the new dataset.
Y-hat procedure -2-

NB:

- the y-hat transformation can also be used to combine a set of indicators into a single stack-able independent variable
  - e.g., define a multiple regression with utilities as dependent variable and as independents, e.g., occupation, income, autonomy in work, etc. in order to derive a single y-hat for job-status

- The y-hats contain exactly the same explanatory information as the original independent variable(s) as they are nothing else than a linear transformation of the original variable(s).

Empirical example -1-

We will use for illustration a subset from the EES 2004, which is a subset of variables and of cases (England only). We’ll work with the following variables (see the codebook of EES04 for full question texts etc.):

- Respondent ID
- Political interest score (q025)
- Electoral utility items for 4 parties (q11a-d)
- Left/right self-placement of respondent and
- Respondents’ perceptions of left/right positions of 4 parties (in the same order as above) (q19 & q19a-d)
- EU integration stance of respondent and perceived stance of parties (q13 & q13a-d)
- Sex (d03)
- Class (d07)
A stacked dataset will be made with the stacked utility items as dependent variable and stacked left/right distances as independent.

The following syntax creates identifiers necessary for stacking four parties (as many are required as there are parties to be stacked):

```
compute p1=1.
compute p2=2.
compute p3=3.
compute p4=4.
execute.
```

Compute left/right distances as follows:

```
compute d_LR_lab = abs(q19 - q19a).
compute d_LR_cons = abs(q19 - q19b).
compute d_LR_lib = abs(q19 - q19c).
compute d_LR_UKIP = abs(q19 - q19d).
execute.
```
Empirical example -3-

SPSS menu tabs: data > restructure. This brings you in a wizard:

- 1\textsuperscript{st} step choose a kind of restructuring. Choose the first option (restructure selected variables into cases)

- 2\textsuperscript{nd} step: define the number of variable groups, this is the number of stacked variables that will be created in the new datafile, each from a number of separate variables in the unstructured file.
  - In our example, the number is 5: (a) identification of parties, (b) utilities of the parties, (c) EU distances, (d) left/right distances, and (e) class.
  - NB: Sex and Political interest will be included as ‘fixed’ variables, i.e. they will be the same for all (respondent × party) combinations pertaining to the same respondent
Empirical example -4-

3rd step: define the variables that have to be stacked, and define their name in the stacked datamatrix. For example:

- Name 1st target variable ‘utility’ and define PTV variables as the variables from which it will be constructed
- Name the 2nd variable lr_dist, and define LR absolute distance variables as the variables from which it will be constructed
- Name the 3rd variable id_pty, and define p1 to p6 as the variables from which it will be constructed
- Similar for EU distances and class
4th step of the SPSS-wizard involves the creation of ‘index variables’, which is the same as creating identifiers for the stacks. You have already done this by creating the variables p1 to p4, so you may specify ‘none’ (alternatively, if you had not already made the identifier, you can specify in this step the option of 1 index variable).
5th step asks what to do with the variables that are not to be stacked, and what to do with missing data

- When specifying ‘keep’ for the non-selected variables, their values are replicated for all new records that pertain to the same respondent (as we will do for sex and political interest)

- In the 2nd box choose ‘create a case’, as otherwise the resulting file becomes exceedingly non-transparent

- Next step asks whether you want to restructure or to save syntax. In the latter case you have to execute the saved syntax from the syntax window

In the data editor view of SPSS you now find the desired stacked data
Stacking the data in STATA

For Stata users we have a full set of utilities, packaged as ‘ptvtools’, which can be ‘net installed’ into Stata where they become commands like any others. These will eventually be available from the Stata website, but for now you need to type...


- Send an email to Lorenzo.desio@eui.eu if you want to be informed of updates and where these updates can be obtained.
Calls on ptvtools…

- `gendist lrp1-lrp10, respondent(VAR202) impute(mean) round`
  `{impute()} provides one option for dealing with missing data}`
- `ptvstack ptv lrp, contextvars(cid) replace`
  `{ in the variable list each of the variables listed are shifted
  into the correct position in the stacked data. Other variables
  are copied (duplicated) across all the cases belonging to a
  specific respondent }`
- `genyhat ywork: q39_p work_* || yvote: q39_p q25`
  `{ in each variable list the variable to be stacked is listed first.
  The dependent variable is implicit – by default ‘ptv’}`
- `help ptvtools`  `{ will open a viewer that lists the available tools
  and provides links to help text for each one }`
References


