Measuring public support for European integration using a Bayesian IRT model

Michele Scotto di Vettimo

Department of Political Economy
King’s College London

September, 2020
Introduction

In the recent years the European Union has experienced a growing politicisation and contestation.

Nowadays, public attitudes are central to:

- Theories of European integration (Hooghe & Marks 2009)
- Studies of national-level party strategies (Hutter & Grande 2014, Reh et al. 2020, Rauh et al. 2020)

The paper addresses the problem of the measurement of public support for the EU.
I adopt the two-dimensional conceptualisation of EU support put forward by De Vries (2018), which focuses on both the *procedural* and the *substantive* dimension of public support.
Existing measurement approaches

Current approaches to the measurement of public attitudes towards the EU can be divided into two camps:
Current approaches to the measurement of public attitudes towards the EU can be divided into two camps:

- The use of available single-question indicators:
  - Two paradigmatic examples could be the Eurobarometer “membership” question or the “preferred speed of integration” one.
Existing measurement approaches

Current approaches to the measurement of public attitudes towards the EU can be divided into two camps:

- The use of available single-question indicators:
  - Two paradigmatic examples could be the Eurobarometer “membership” question or the “preferred speed of integration” one.

- The combination of multiple indicators with dimension-reduction techniques.
  - Example: The Dyad Ratios algorithm (Stimson 1991). Similar to principal component analysis.
Limitations of existing approaches

- Single-question indicators:
  - Poor in accounting for possible multidimensionality of EU support
  - Very few indicators have been measured over long time periods
  - Data series have interruptions or question wording might change

- Dyad Ratios algorithm:
  - Lacks an individual-level model of response
  - Cannot treat indicators with neutral or ordinal responses
  - Estimated series has an artificial metric.
IRT models

Item response models were developed in psychometric theory to measure latent attributes (e.g., ideology or ability) of individual respondents. For instance, if you've ever taken a GMAT or GRE test, you've been administered an IRT-based test! Individual-level IRT uses responses to different items to calculate respondent’s “knowledge” (the attribute $\theta$), by taking into account how difficult a given question is and how well it discriminates between “knowledgeable” and “unknowledgeable” respondents.
IRT models

- Item response models were developed in psychometric theory to measure latent attributes (e.g., ideology or ability) of individual respondents.
IRT models

- Item response models were developed in psychometric theory to measure latent attributes (e.g., ideology or ability) of individual respondents.

- For instance, if you’ve ever taken a GMAT or GRE test, you’ve been administered an IRT-based test!
IRT models

- Item response models were developed in psychometric theory to measure latent attributes (e.g., ideology or ability) of individual respondents.

- For instance, if you’ve ever taken a GMAT or GRE test, you’ve been administered an IRT-based test!

- Individual-level IRT uses responses to different items to calculate respondent’s “knowledge” (the attribute $\theta$), by taking into account how difficult a given question is and how well it discriminates between “knowledgeable” and “unknowledgeable” respondents.
Bayesian IRT with aggregated data

IRT models can be adapted to deal with population-level data.
Bayesian IRT with aggregated data

IRT models can be adapted to deal with population-level data.

- Imagine that:
  - We know that latent support in the population is normally distributed with mean $\mu$ and standard deviation $\sigma$, so that we can calculate the probability of drawing an individual with a certain amount of latent support.
  - We know how difficult and how discriminatory question $q$ is.
  - The probability of answering a particular question in a pro-European way is a function of respondent’s latent support and question’s difficulty and discrimination.
Bayesian IRT with aggregated data

IRT models can be adapted to deal with population-level data.

- Imagine that:
  - We know that latent support in the population is normally distributed with mean $\mu$ and standard deviation $\sigma$, so that we can calculate the probability of drawing an individual with a certain amount of latent support.
  - We know how difficult and how discriminatory question $q$ is.
  - The probability of answering a particular question in a pro-European way is a function of respondent’s latent support and question’s difficulty and discrimination.

- Hence:
  - We could calculate the expected number of pro-European responses observed if question $q$ is administered to a sample of $N$ individuals randomly drawn from the population.
Bayesian IRT with aggregated data

In fact, we find ourselves in the opposite situation:

- The number of pro-European responses given to question $q$ is a known quantity (share $\times$ sample size)
- Yet, we do not know how difficult and discriminatory question $q$ is
- Nor we know the average latent support of the population
Therefore:

- We write a model (a system of equations) that formalises the relationship between the different known (observed data) and unknown (model parameters) quantities.

- We use a Bayesian programme to “solve” our model and find the most likely combination of model parameters that generates the data that we actually observe.
Bayesian IRT addresses different shortcomings of existing approaches.

With regard to single-question indicators, IRT models:
- Can draw from multiple items to better deal with multidimensionality
- Can deal with interruptions in the administration of a specific item

Similarly, compared to the DR algorithm, IRT models:
- Starts from an individual-level model of response
- Accomodate questions with neutral or ordinal answer options
To apply Bayesian IRT to the estimation of EU support, I identified 121 Eurobarometer questions asked between 1973 and 2019.

To be included, a question had to be asked at least 4 times and:

- Tap on aspects related to EU trust, attachment, evaluation of EU membership (procedural dimension)
- Or, ask about one of the following: preferences for the level of government responsible for a policy, whether more or less EU action or EU-level harmonisation should take place or whether a common EU policy should be developed in a given domain (substantive dimension)
Application to public support for the EU

- Overall, from 591 to 1936 question-semester dyads are available for each EU member state
- The data: share of pro-EU and neutral responses, a question identifier and the fieldwork date for each item administration
- The model is an adaptation of McGann et al. (2019) and it is implementend using the software JAGS.
- Question parameters are constrained to be the same for all countries, so as to ensure cross-country comparability.
Estimated series I

Belgium

France

Germany

Italy

Luxembourg

Netherlands

Denmark

Ireland

United Kingdom

Greece

Spain

Portugal

Michele Scotto di Vettimo

Measuring public support for EU using Bayesian IRT
Estimated series II

Michele Scotto di Vettimo
Measuring public support for EU using Bayesian IRT
Convergent validity

Is the IRT measure associated with other valid measures of EU support?
Convergent validity

Comparison with Anderson & Hecht (2018) “preference for Europe” measure estimated with the DR algorithm for four member states.

France  
\[ r = .75 \]

Germany  
\[ r = .69 \]

Italy  
\[ r = .93 \]

United Kingdom  
\[ r = .83 \]
Convergent validity

Comparison with Guinaudeau & Schnatterer (2019) “EU mood” measure estimated with the DR algorithm for all EU member states.

European Union

\[ r = .76 \]
Model fit

If we accept the IRT estimates to be a good measure of latent public EU support, how well can we predict the true share of positive responses recorded for the 121 items?

<table>
<thead>
<tr>
<th></th>
<th>Root mean sq. error</th>
<th>By-item R-sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRT model</td>
<td>3.39</td>
<td>0.384</td>
</tr>
<tr>
<td>DR algorithm</td>
<td>3.56</td>
<td>0.319</td>
</tr>
<tr>
<td>“EU mood”</td>
<td>3.61</td>
<td>0.301</td>
</tr>
<tr>
<td>“Membership” question</td>
<td>3.78</td>
<td>0.229</td>
</tr>
<tr>
<td>Item means only</td>
<td>4.31</td>
<td>0 (by construction)</td>
</tr>
</tbody>
</table>
Construct validity

Is the measure associated with other theoretically related concepts?

Namely, is the estimated EU support negatively associated with vote share of Eurosceptic parties?
## Construct validity

Tobit regression models of public support for Europe and Eurosceptic vote share.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV: Eurosceptic vote share</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRT-estimated EU support</td>
<td>-1.239***</td>
<td>-0.670**</td>
</tr>
<tr>
<td>Previous Eurosceptic vote share</td>
<td>0.286*</td>
<td>(0.111)</td>
</tr>
<tr>
<td>Election type (1=European)</td>
<td>2.731</td>
<td>(1.523)</td>
</tr>
<tr>
<td>% exclusive national identity</td>
<td>-0.116</td>
<td>(0.146)</td>
</tr>
<tr>
<td>% foreign-born population</td>
<td>-0.896</td>
<td>(0.904)</td>
</tr>
<tr>
<td>% foreign-born labour force</td>
<td>-0.0865</td>
<td>(0.877)</td>
</tr>
<tr>
<td>Net contribution to EU budget</td>
<td>-0.350</td>
<td>(0.461)</td>
</tr>
<tr>
<td>Annual GDP growth rate</td>
<td>0.214</td>
<td>(0.183)</td>
</tr>
<tr>
<td>Annual unemployment rate</td>
<td>0.0609</td>
<td>(0.262)</td>
</tr>
<tr>
<td>Year</td>
<td>0.877***</td>
<td>(0.247)</td>
</tr>
<tr>
<td>Constant</td>
<td>66.33***</td>
<td>-1708.9***</td>
</tr>
<tr>
<td>Observations</td>
<td>259</td>
<td>259</td>
</tr>
<tr>
<td>Country fixed effects</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>LR $\chi^2$</td>
<td>48.82</td>
<td>255.14</td>
</tr>
<tr>
<td>Prob. $&gt;\chi^2$</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-507.552</td>
<td>-404.389</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 
Conclusion and discussion

Bayesian IRT can produce valid measures of EU support from population-level data:
Bayesian IRT can produce valid measures of EU support from population-level data:

- IRT models provide a more precise measurement of latent support if compared to available alternative techniques.
Conclusion and discussion

Bayesian IRT can produce valid measures of EU support from population-level data:

- IRT models provide a more precise measurement of latent support if compared to available alternative techniques
- They are grounded in an individual-level model of response
Bayesian IRT can produce valid measures of EU support from population-level data:

- IRT models provide a more precise measurement of latent support if compared to available alternative techniques.
- They are grounded in an individual-level model of response.
- They allow for a sounder treatment of neutral responses.
IRT models can be used to estimate EU support also where comparable measures do not exist:

- Selecting a subset of questions related to a specific policy area, it is possible to estimate public support for EU integration in that domain.
References I


Thank you all for your comments!