KANTAR PUBLIC

Dealing with respondents that do not express a firm preference in political opinion polls

Luke Taylor

ESRA Conference 2021

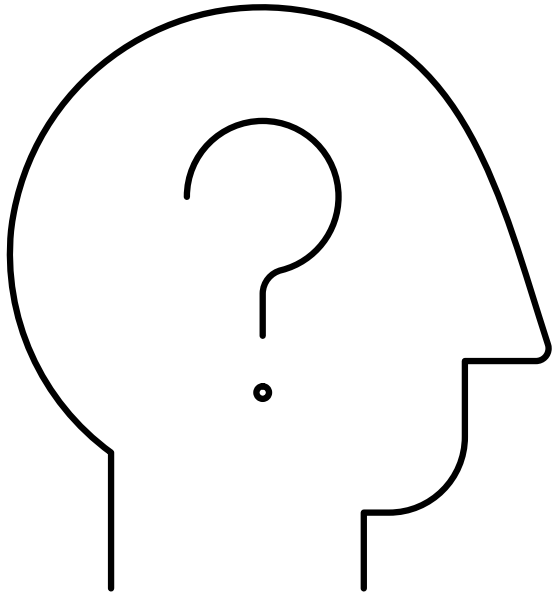
9.7.2021

Background and approach

Background

A major challenge when conducting political polling is deciding how to deal with respondents that do not express a firm voting preference
(that are unsure or that prefer not to say who they will vote for)

This paper outlines the methodology used in the UK by Kantar to estimate the voting intention of individuals that do not express a firm preference. We will also use the 2019 General Election to evaluate the success of the approach.



What is the risk of bias?

Item non-response bias depends on two factors*:

$$B(\bar{Y}_r) = Wm(\bar{Y}_r - \bar{Y}_m)$$

Where:

W_m = the item non-response rate

$\bar{Y}_r - \bar{Y}_m$ = the difference between the mean for respondents and the mean for non-respondents

One approach commonly taken by pollsters is to exclude those that do not provide a preference altogether from voting intention estimate. However, this makes the bold assumption that the data is **missing completely at random** ($\bar{Y}_r - \bar{Y}_m = 0$)

**See: Kalton, G., 2020. Introduction to survey sampling (Vol. 35). Sage Publications.*

Summary of the Kantar approach in the UK

Who will you vote for in your local constituency?

Final 2019 GE poll

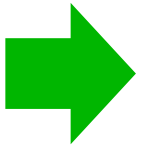
14%

item non-response
(among likely voters)

Source: Kantar final 2019 GE poll (f/w 9th-11th December 2019) , *n=2,815*

Summary of the Kantar approach in the UK

Who will you vote for in your local constituency?



Which parties might you consider voting for?

Final 2019 GE poll
14%
item non-response
(among likely voters)

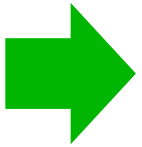
Final 2019 GE poll
9%
item non-response

-5pp

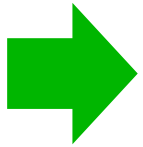
Source: Kantar final 2019 GE poll (f/w 9th-11th December 2019) , *n=2,815*

Summary of the Kantar approach in the UK

Who will you vote for in your local constituency?



Which parties might you consider voting for?



If no information: impute

Final 2019 GE poll
14%
item non-response
(among likely voters)

Final 2019 GE poll
9%
item non-response

-5pp

Final 2019 GE poll
0%
item non-response

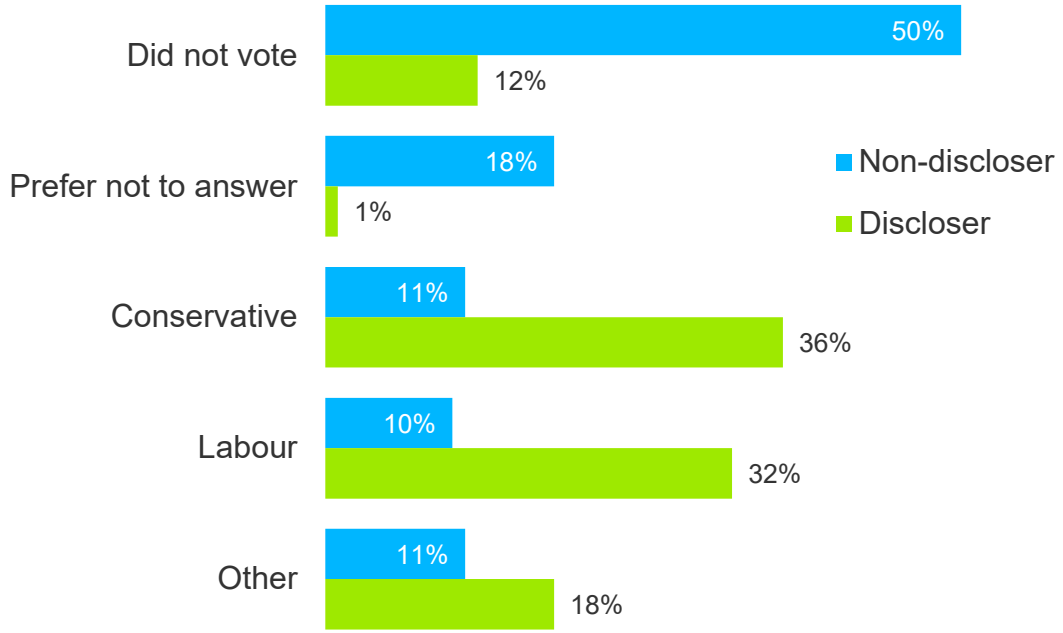
-9pp

Source: Kantar final 2019 GE poll (f/w 9th-11th December 2019) , *n=2,815*

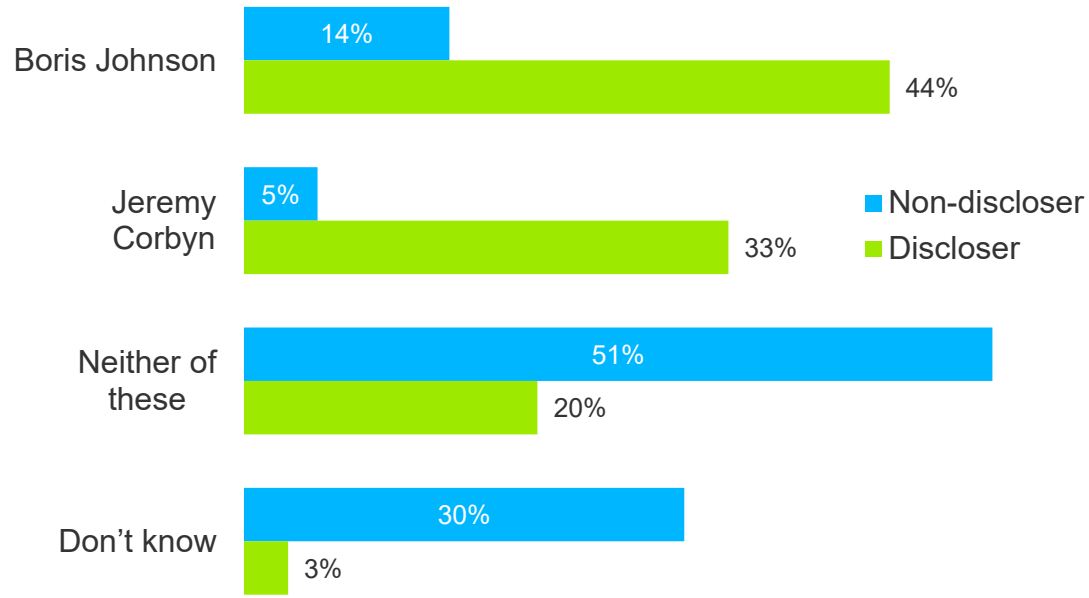
Challenges for imputation – having informative predictors

People that do not disclose a voting intention, are also more likely to provide non-informative responses at other questions

Vote in 2017 General Election



Best leader for Britain



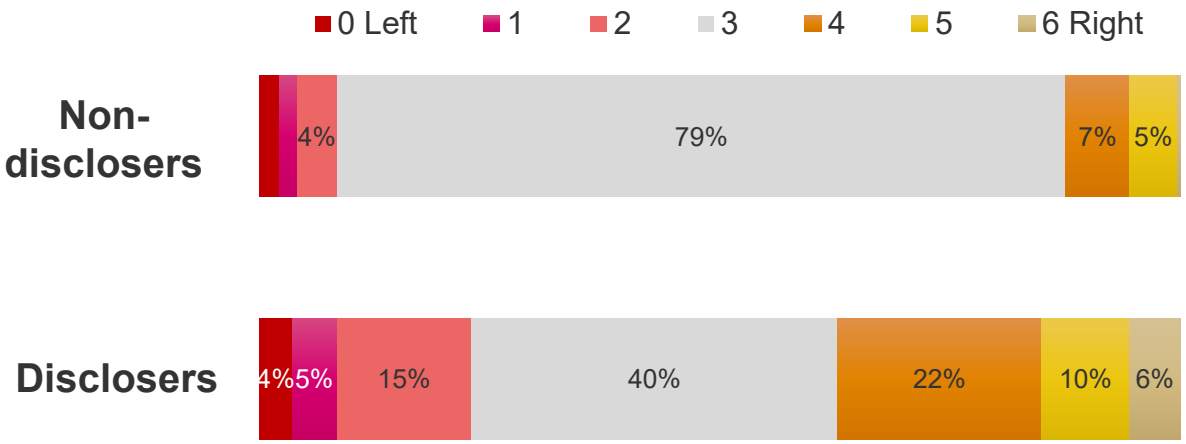
Source: Kantar final 2019 GE poll (f/w 9th-11th December 2019) - unweighted

Base: Non-disclosers (n=273), Disclosers (n=2,542)

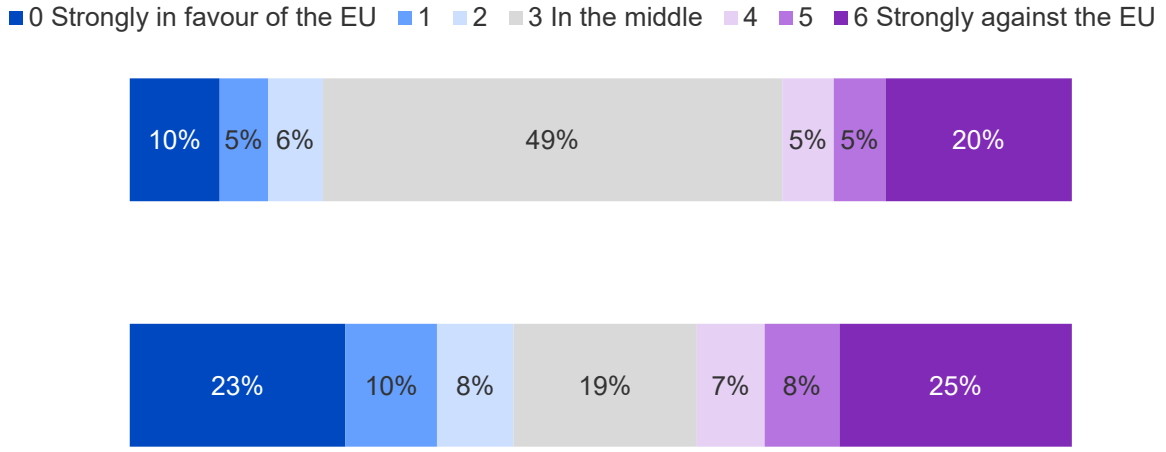
Challenges for imputation – having informative predictors

Our polling questionnaire was designed to include some items which people must give an opinion at. Although this had limited success.

In politics people sometimes talk of Left and Right. Where would you place yourself on the following scale?



Now thinking about the EU. Where would you place yourself on the following scale?



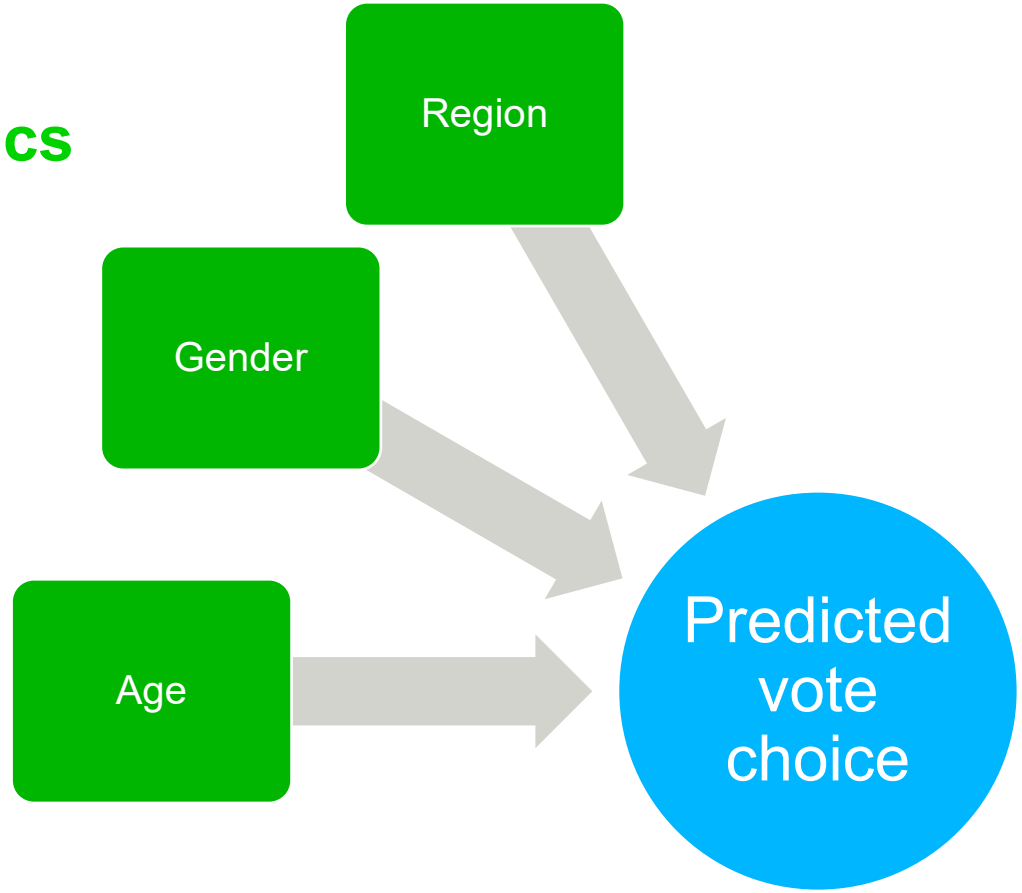
Source: Kantar final 2019 GE poll (f/w 9th-11th December 2019) - unweighted

Base: Non-disclosers (n=273), Disclosers (n=2,542)

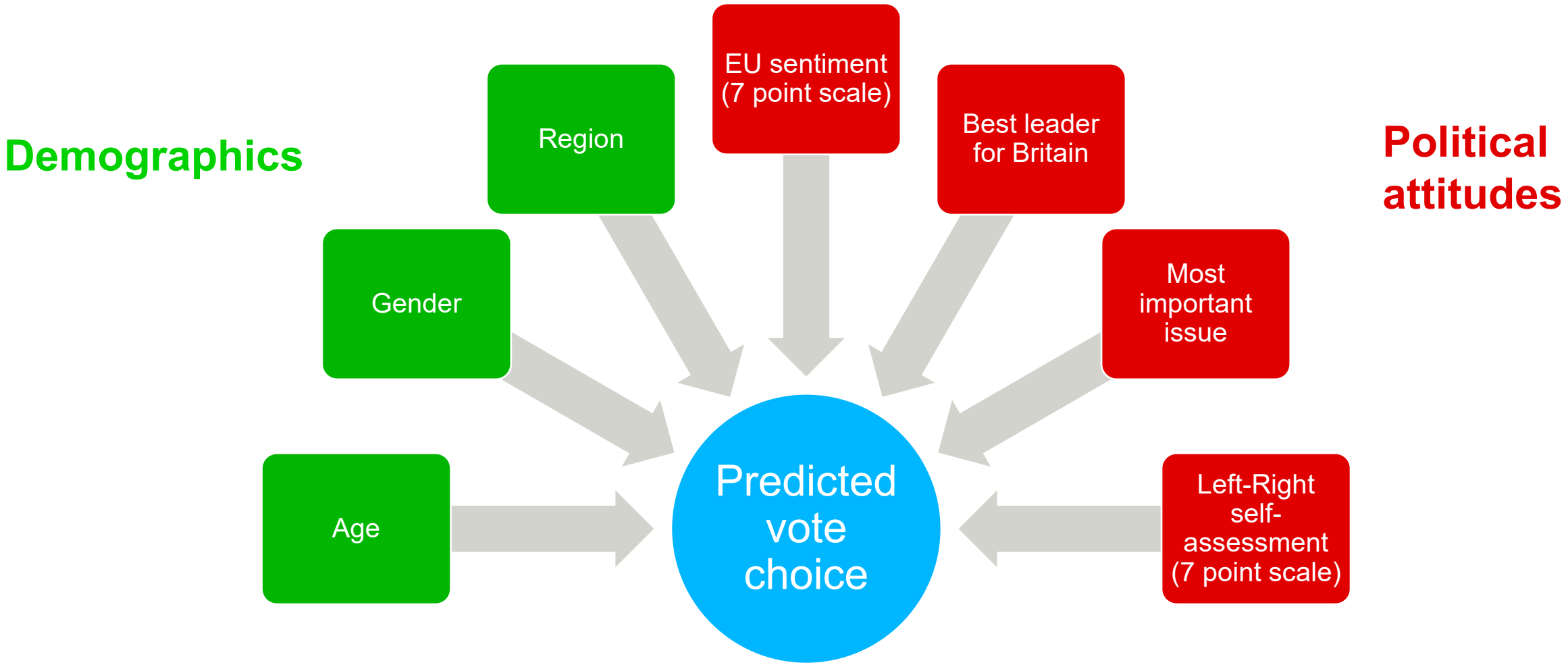
Final predictors used

Final predictors used

Demographics



Final predictors used



Imputation approach

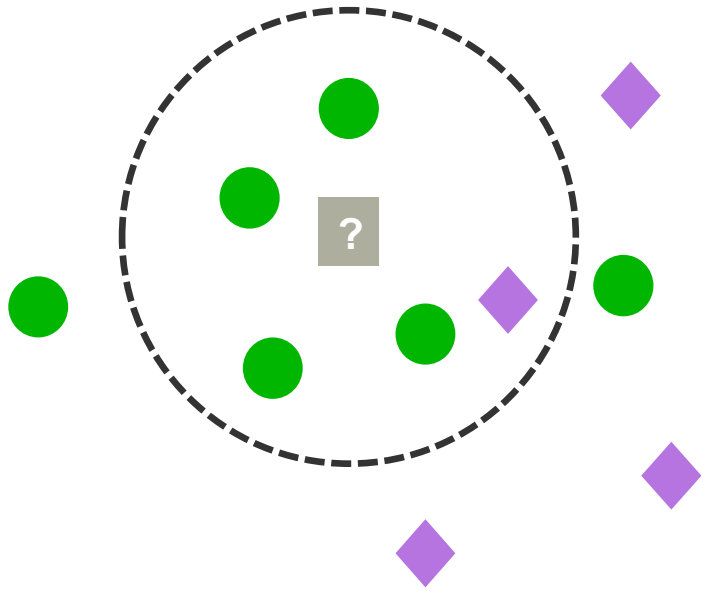
We trialled different imputation approaches, but in the end chose **K-Nearest Neighbours**

Advantages of KNN over other approaches:

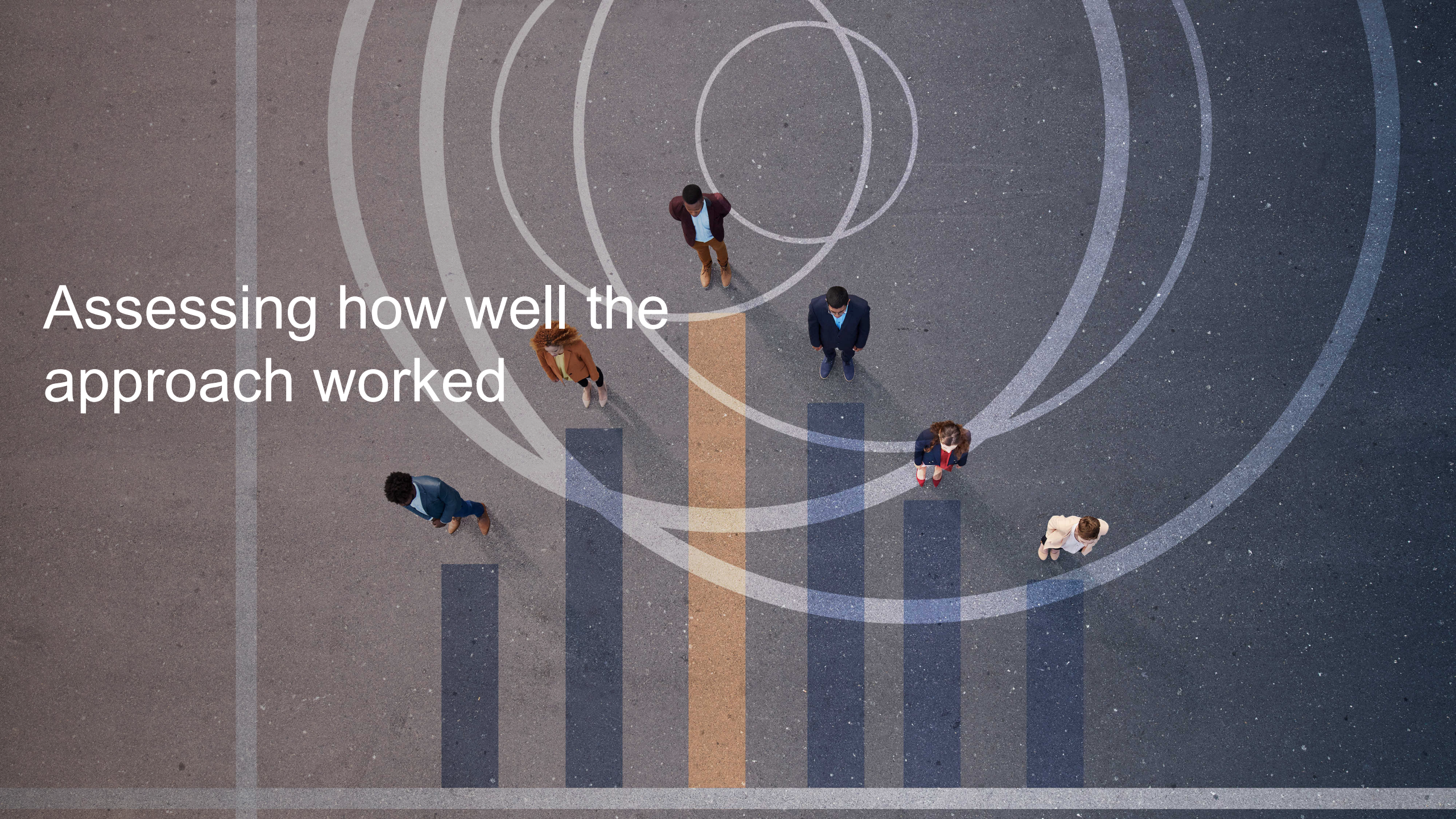
- Non-parametric - no assumption for the underlying data distribution
- Algorithm works well for multi-class problems (where some of the categories are very small).
    - We found alternative approaches ***almost always*** assigned people to the two major parties. Whereas KNN was more likely to assign people to minor parties – which we felt was more realistic
- Easy to implement in our existing data processing system

There are also some potential disadvantages:

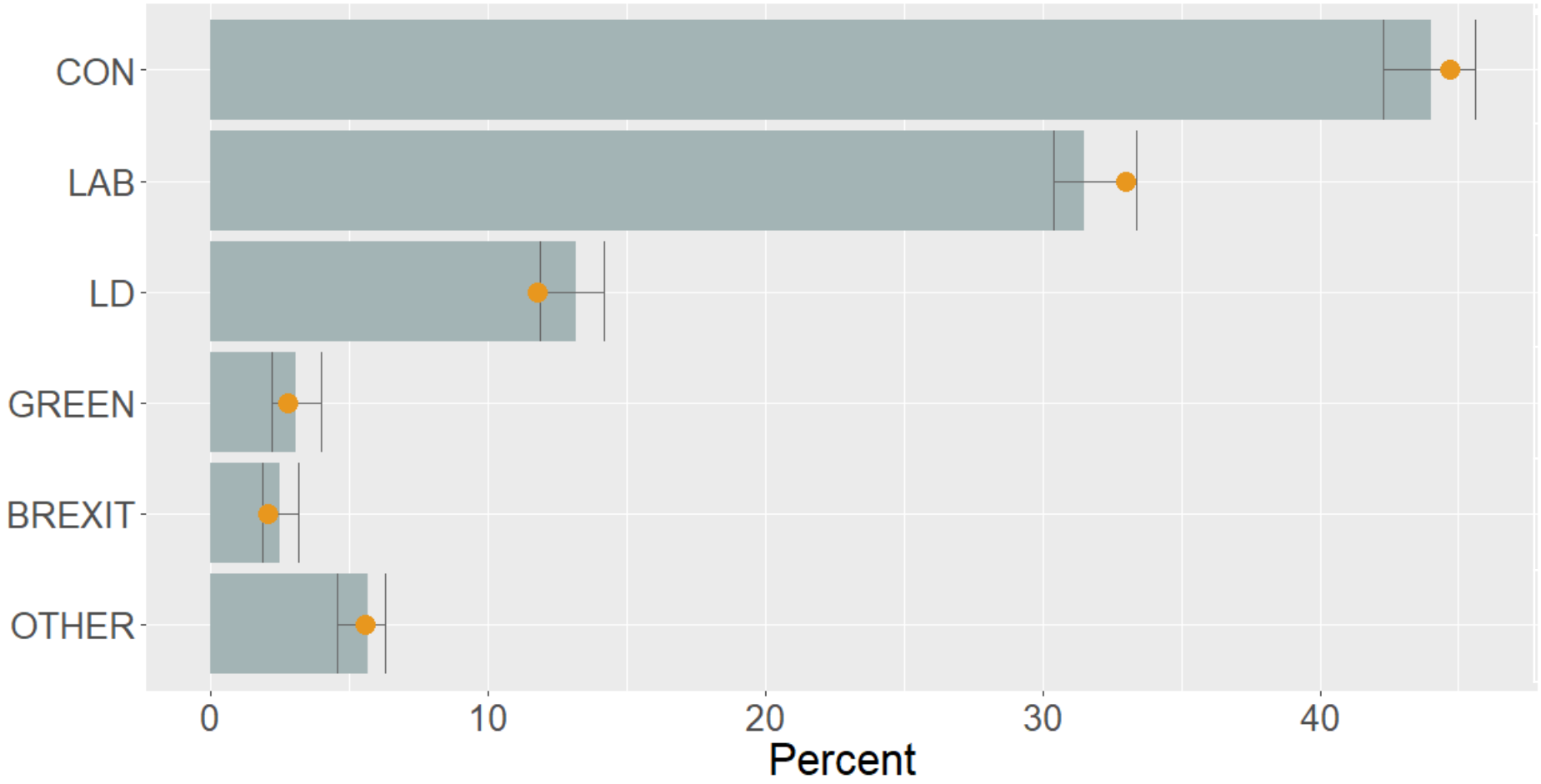
- Sensitive to outliers
- Each feature is assumed to be of the same importance



Assessing how well the approach worked



Final Kantar Voting Intentions vs GB election results

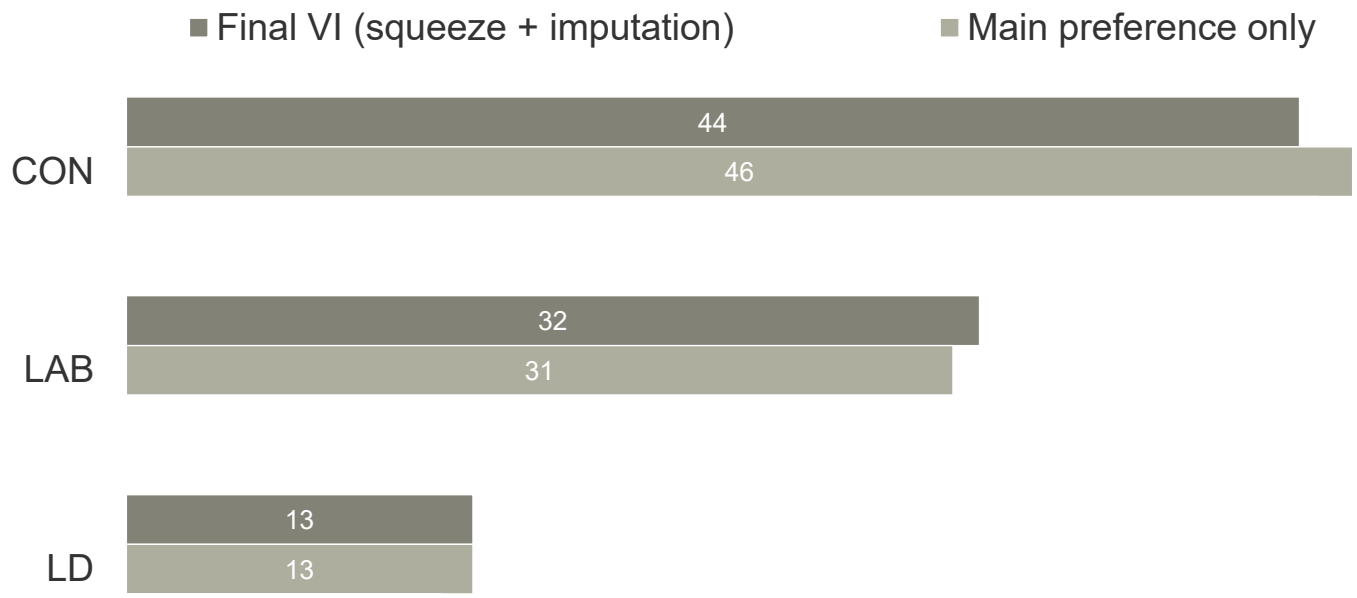


	Poll	Election result
CON	44%	44.7%
LAB	32%	33.0%
LD	13%	11.8%
GREEN	3%	2.8%
BREXIT	3%	2.1%
OTHER	5%	5.6%

Sources: Kantar poll, fieldwork = 9-11th December 2019, n=2,815 – weighted by likelihood to vote
http://www.britishpollingcouncil.org/the-performance-of-the-polls-in-the-2019-general-election//

What would have happened had we just excluded respondents that did not express a firm preference?

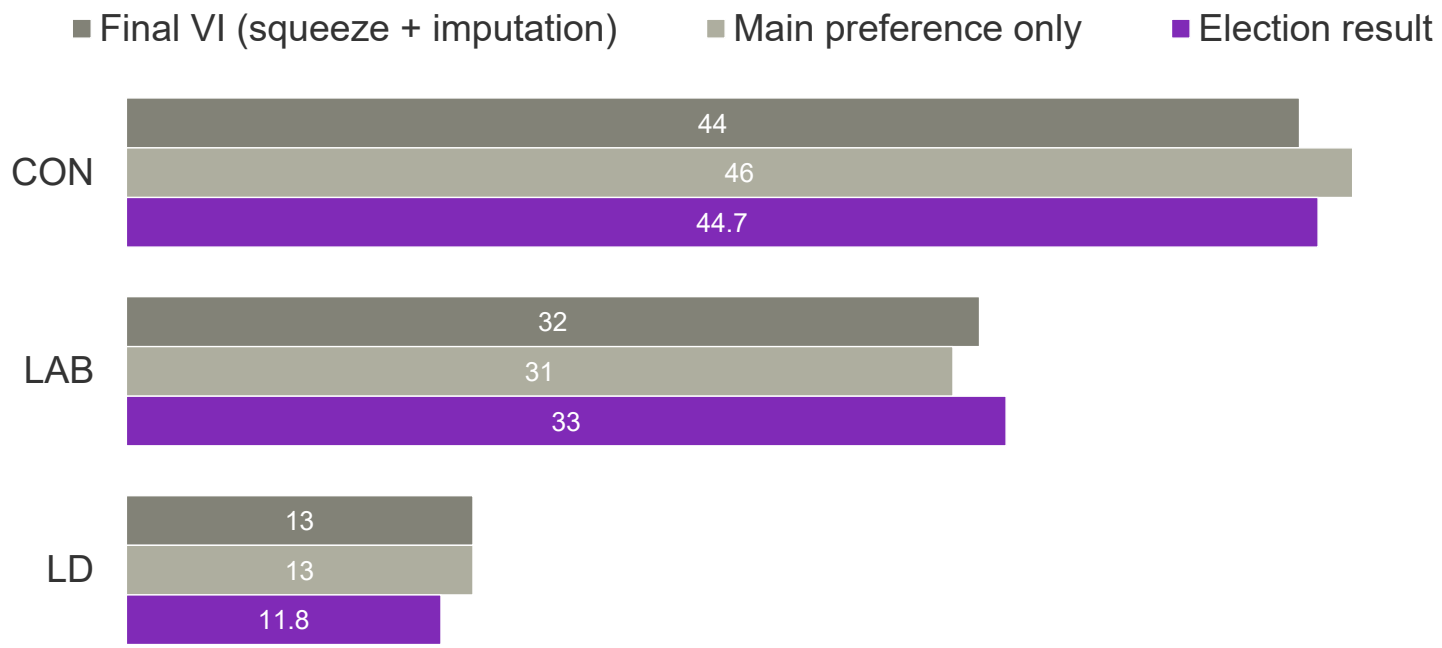
Voting intentions – top three parties



Sources: Kantar poll, fieldwork = 9-11th December 2019, n=2,815 – weighted by likelihood to vote
http://www.britishpollingcouncil.org/the-performance-of-the-polls-in-the-2019-general-election//

What would this have meant for accuracy?

Voting intentions – top three parties



Mean Absolute Error	
Final VI	1%pt
Main preference only	1.5%pt

Sources: Kantar poll, fieldwork = 9-11th December 2019, n=2,815 – weighted by likelihood to vote
http://www.britishpollingcouncil.org/the-performance-of-the-polls-in-the-2019-general-election//

Final remarks

Our approach – using a squeeze question and imputation – improved accuracy of our final poll

However, the impact was very slight

1. Those that do not provide a voting intention are pretty distinctive – likely they do systematically differ from those that disclose a preference
2. But they are only a minority of likely voters, and this limits the potential impact of bias

Imputing data is challenging due to the difficulty in collecting informative covariates from a group of the population that are not very politically engaged

- For next election, try to design new questions to better collect this information
- Questions required likely to vary between elections depending on what the issues driving vote choice are (e.g., view on EU was particularly salient for 2019 GE)

Thank you

luke.taylor@kantar.com