

SUPPLEMENTARY MATERIAL - Estimating Ideal Points of British MPs Through Their Social Media Followership

1. Fitting a Correspondence Model

This appendix outlines exactly how ideal points can be derived from network data using correspondence analysis. CA is a multidimensional scaling technique that is conceptually similar to principal components analysis (PCA) but can be applied to categorical data instead of continuous (Greenacre 2010). The general mathematical process of CA is relatively straightforward (following Barberá et al. (2015)):

Consider an $i \times j$ contingency table - in this case, our \mathbf{Y} follower adjacency matrix - with elements y_{ij} and total number of observations $e = \sum_{ij} y_{ij}$:

Step 1: Convert \mathbf{Y} into a correspondence matrix \mathbf{P} by dividing \mathbf{Y} by its total sum: $\mathbf{P} = \mathbf{Y} / \sum_{ij} y_{ij}$. This converts all elements of \mathbf{Y} to proportions.

Step 2: Calculate row and column *masses* of \mathbf{P} by summing the elements of each individual row and column: $r_i = \sum_j p_{ij}$ and $c_j = \sum_i p_{ij}$. These are then used to construct diagonal matrices of $\mathbf{D}_r = \text{diag}(\mathbf{r})$ and $\mathbf{D}_c = \text{diag}(\mathbf{c})$.

Step 3: Compute a matrix of standardised residuals \mathbf{S} where $\mathbf{S} = \mathbf{D}_r^{1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}^T)\mathbf{D}_c^{1/2}$. These residuals reflect the difference between expected and observed values in each element of \mathbf{P} based on their corresponding r_i and c_j . By standardising them in this way, CA is able to control for α_i (political interest of user i) and β_j (popularity of actor j) effects by re-weighting rows and columns based on how populated they are. This similarly helps to adjust for case where sample sizes are smaller (i.e. low follower counts).

Step 4: Subsequently, each element of matrix \mathbf{S} contains a residual S_{ij} which is treated by the CA model as a reflection of the ideal point distance between user i (Θ_i) and actor j (Φ_j) in the latent ideological space. The singular value decomposition (SVD) of \mathbf{S} is then calculated, such that $\mathbf{S} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^T$ where $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$. SVD is the primary algorithmic procedure for identifying the dimensional space, onto which the column and row coordinates can be projected: $\Psi = \mathbf{D}_r^{1/2}\mathbf{U}$ for rows (ordinary users) and $\Gamma = \mathbf{D}_c^{1/2}\mathbf{V}$ for columns (MPs). In this case, these coordinates for both the rows and columns reflect their ideal points in the latent ideological space.

2. Sample Filter Robustness Checks

This appendix consists of a robustness check undertaken to assess the performance of the correspondence model using different sample subsets. As is explained in the data collection section of the main paper, the set of ordinary follower profiles is filtered to only include those who follow at least 10 MPs. This is to ensure that an especially informative subset of users is used to estimate ideal points of MPs, reducing noise and optimising model performance. The choice of setting the sample filter at a minimum of 10 was informed by Barberá's (2015) original paper but in order to test the robustness of the model estimates in relation to this threshold, the CA model was ran using subsets of the follower

data at multiple different thresholds. The ideal points of MPs generated by the model at each of these thresholds was validated against the same expert survey estimates to assess how well the model performed at each threshold, and how much the estimates varied in relation to this filtering. Model validation was conducted using Pearson's r correlation coefficients of the model estimates against the expert validation estimates for both the overall model (between-party accuracy) and for Conservative and Labour MPs respectively (within-party accuracy). 20 iterations of the correspondence model were ran, increasing the sample filter threshold at each point starting at all users who followed at least 1 MP (all profiles) increasing by increments of 1 up to 5, then by increments of 5 up to 20, then increments of 10 up to 50, and finally increments of 50 up to 500. The details of the sample threshold for each iteration can be found in **Table 1**:

Table 1. Details of each sample subset filtering at different MP following thresholds. Between r shows the Pearson's correlation coefficient of the overall model estimates against the expert validation estimates. CON r and LAB r illustrate the within-party correlation coefficients.

Iteration	Threshold	MPs	Users	Connections	Between r	CON r	LAB r
1	1	591	4,460,657	20,048,554	0.83	0.06	0.78
2	2	591	2,161,097	17,748,994	0.97	0.88	0.86
3	3	591	1,450,688	16,328,176	0.97	0.91	0.86
4	4	591	1,094,788	15,260,476	0.97	0.90	0.86
5	5	591	878,815	14,396,584	0.97	0.89	0.85
6	10	591	424,297	11,443,165	0.97	0.84	0.81
7	15	591	260,849	9,532,641	0.96	0.78	0.77
8	20	591	178,909	8,158,385	0.96	0.73	0.74
9	30	591	99,736	6,274,876	0.95	0.66	0.68
10	40	591	62,735	5,017,414	0.95	0.61	0.64
11	50	591	42,351	4,120,608	0.94	0.58	0.59
12	100	591	12,079	2,098,538	0.94	0.51	0.47
13	150	590	5,496	1,309,591	0.93	0.49	0.46
14	200	590	2,856	857,265	0.93	0.43	0.56
15	250	590	1,721	606,195	0.91	0.27	0.44
16	300	590	1,080	430,673	0.68	0.13	0.16
17	350	590	694	306,490	0.06	0.07	0.35
18	400	590	434	209,697	0.19	0.04	0.45
19	450	590	278	144,063	0.05	0.09	0.43
20	500	590	169	92,439	0.20	0.01	0.27

Using the entire sample of followers with a minimum threshold of MPs followed = 1, the model performs relatively poorly, particularly when discriminating between the ideal points of Conservative MPs. There is a sizeable improvement in both between and within-party accuracy when increasing the minimum threshold to 2, with between-party accuracy remaining reasonably stable until iteration 16 when it declines dramatically. After iteration 5, the within-party accuracy of the model begins to decline for both Conservative and Labour MPs. This trend is intuitive given that one would expect there to be an optimal "Goldilocks" point when subsetting the follower data to include only the most informative users. Setting the threshold too low will include users who are not especially politically informed and thus have less ability to discriminate ideologically between MPs, and setting

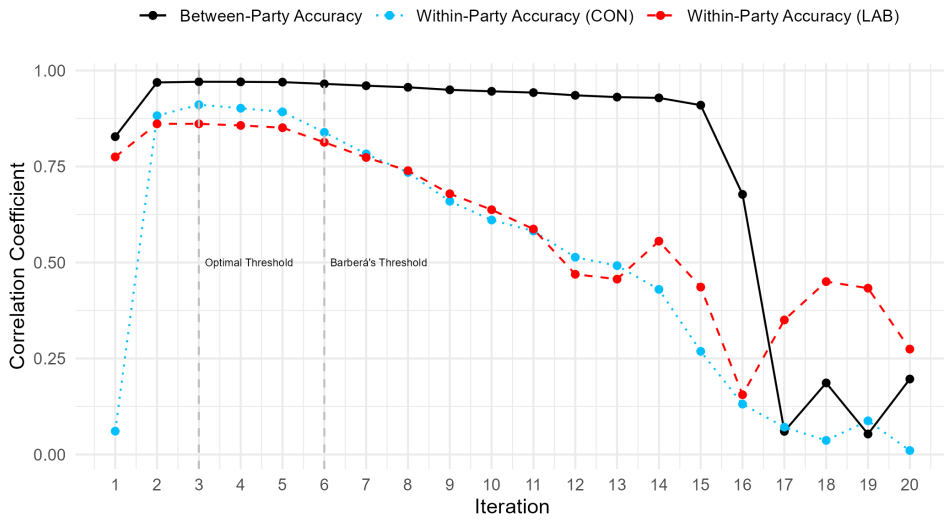


Figure 1. Correlation coefficient represents the Pearson's r between the model estimates and the expert validation estimates at each iteration sample threshold. Optimal model performance is achieved in iteration 3 at a threshold of 3 and Barberá's threshold of 10 was set in iteration 6.

the threshold too high will only include users who follow a significant proportion of MPs from across the spectrum, flattening out the ideological component altogether. In this case, the model estimates are most closely correlated with the expert estimates (both between and within) when setting the threshold at a minimum of 3 MPs, notably lower than Barberá's original threshold of 10. However, there are also efficiency considerations to take into account. At lower thresholds, the follower adjacency matrices are far larger and, as such, dimensionality reduction can be far more computationally intensive and time consuming. Thus, one could justifiably trade-off some degree of model accuracy in favour of a less time consuming and computationally demanding scaling procedure.

3. CA Model Dimensions

This appendix contains details of the multiple dimensions generated from the correspondence analysis conducted on the follower adjacency matrix.

3.1 Variance Capture By Dimension

Figure 2 illustrates the individual variance capture of each of the first 10 dimensions as a Scree plot:

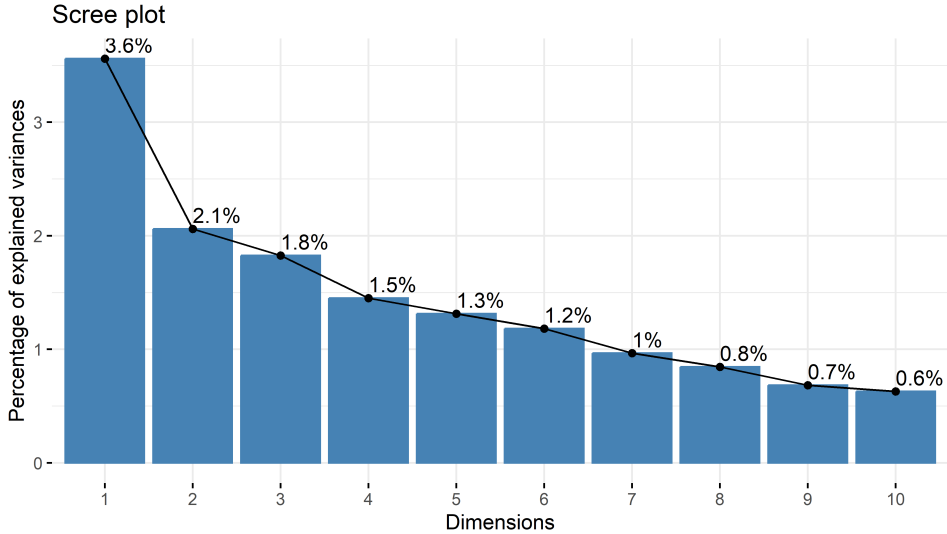


Figure 2. Scree plot demonstrating the percentage of explained variance captured by each of the first 10 dimensions of the CA model.

Dimension 1 captures the largest amount of the variance in the Twitter follower adjacency matrix, although this overall proportion is still low (3.6%). There is a steep drop in the amount of variance capture in dimension 2 (2.1%) which gradually reduces further for each additional dimension in the model. Although the first dimension has been shown in the paper to approximate an ideological component, the fact that it only accounts for 3.6% of the overall variance in the follower matrix indicates that much of the variance in the data cannot be solely explained by ideology. Nevertheless, using only the first dimension in the model still produces a highly accurate measure of MP ideal points.

3.2 *Treating Nationalist MPs as Supplementary Columns*

When conducting correspondence analysis on the follower adjacency matrix, it is explained in the paper that MPs representing national parties are initially excluded from the scaling process and treated as supplementary columns. Consequently, users who exclusively follow national MPs are also excluded from the initial scaling procedure and treated as supplementary rows. These supplementary columns and rows in the matrix and then retroactively projected onto the dimensional space after it has been scaled to ensure that the component being scaled is ideology. This is because the regional component to nationalist parties overwhelms their ideological one. **Figure 3** illustrates a 2-dimensional scatterplot comparing the original model estimates when treating the nationalist MPs as supplementary columns against model estimates when including all MPs in the scaling procedure.

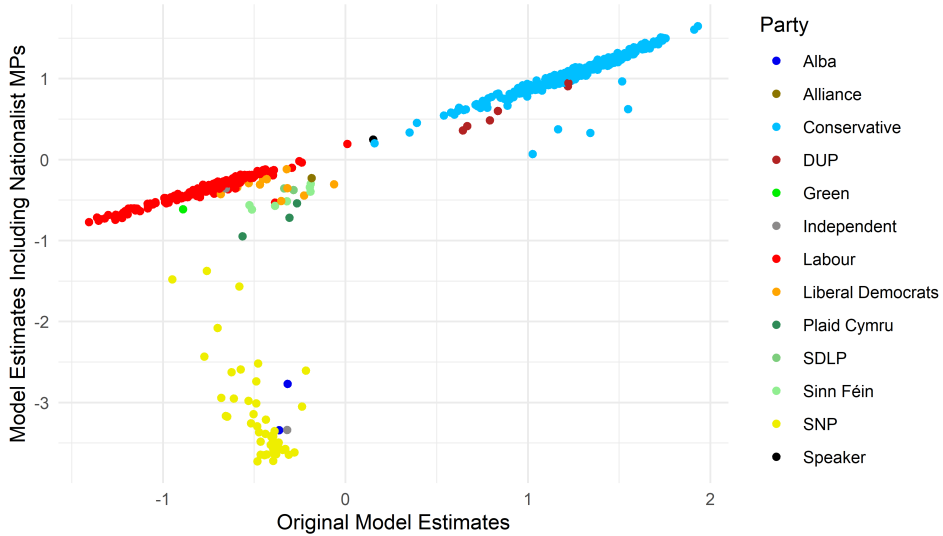


Figure 3. Original CA Model Estimates vs. Model Estimates Including National MPs During Scaling

The estimates for non-nationalist MPs stay approximately equal in both models, but the estimates for nationalist MPs are skewed significantly. This is particularly the case for SNP and Alba Party MPs. There is also a notable skew for MPs of non-nationalist parties who represent constituencies based outside of England. This suggests that including nationalist parties in the scaling process and not treating them as supplementary columns overwhelms the ideological component captured by the first dimension of the model, where nationalist MPs and MPs representing non-English constituencies will instead cluster around a nationalist element. In order to remove this effect, it is subsequently justifiable to treat nationalist MPs as supplementary columns in the initial scaling process and retroactively project them onto the dimensional space.

3.3 Interpreting Additional Dimensions

It has been confirmed in the main paper that the first dimension of the CA follower model approximates left/right ideology. It is uncertain what components the additional model dimensions beyond the first capture, pending deeper investigation. **Figure 4** illustrates the first two dimensions of the CA follower model plotted on a 2-dimensional scatterplot:

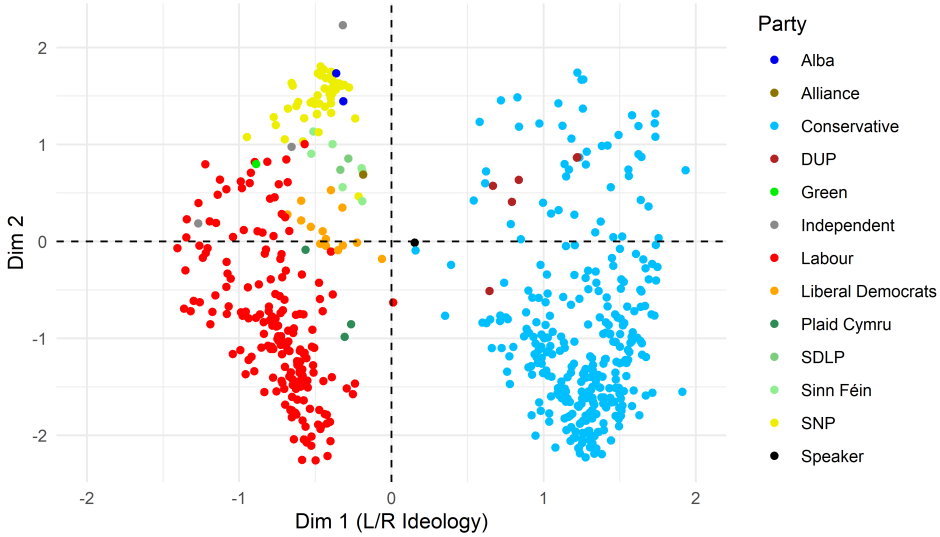


Figure 4. CA Model Dimension 1 Estimates vs. CA Model Dimension 2 Estimates

It is clear that the second model dimension cuts across partisan and ideological boundaries, although nationalist party MPs do still appear to cluster relatively closely together. Initial investigation into the second and third dimensions have so far proven difficult to understand, and further analysis will be needed to ascertain what these components capture. For the purpose of this paper, interpretation of these additional dimensions is not essential given that the first dimension provides an adequate approximation of left/right ideology. It is important to note that these additional dimensions do not necessarily represent additional dimensions of ideology *per se*; rather, they capture whatever additional dimensions drive the follower networks of MPs. An intuitive assumption might be a geographical component to MP follower networks, where ordinary users are more likely to follow MPs who represent constituencies closer to where they live, or a seniority component which divides MPs between more senior ministers and party backbenchers. One possibility for identifying what these additional dimensions represent may be to conduct cluster analysis of MPs using their first three dimensional coordinates and inspect the underlying characteristics that MPs within the same clusters share. This way, a more comprehensive understanding of the primary drivers of MP follower networks can be established, incorporating additional dimensions beyond the initial left/right ideological component. This analysis would be complex and beyond the requirements of this paper, but can encourage future work that can build on this initial research.

4. Expert Survey Format & Responses

A short survey was circulated to 133 experts in British politics and political science on the 08/11/2022. The survey was generated through Qualtrics and made accessible to participants via a link attached to an email that was distributed to each of them individually. All responses were anonymous. The experts chosen to participate in the survey were all academics with substantive research expertise in electoral, legislative, or parliamentary politics and/or public opinion with a specific focus on the U.K. In order to source these academics and their contact information, the politics faculty websites of all U.K universities who are either members of the Russell Group and/or ranked in the top 30 U.K universities for politics in any of the Times Higher Education World University Rankings, the Guardian University Guide, or the Complete University Guide were used. Although this sampling method is slightly restrictive, it helped to narrow down the process as efficiently as possible and provide a straightforward way of sourcing an appropriate number of individuals who could be justifiably considered credible sources of expert opinion on British politics. Of the 133 experts contacted and invited to participate, 70 participants completed the survey in full indicating a 53% response rate.

The survey asked participants to place a sub-sample of 30 U.K MPs, 12 U.K political parties, and 13 U.K media organisations on an 11-point scale between 0-10 based on where they believed each MP/party/organisation sat on the left/right ideological spectrum. Following the same format and questioning as the British Election Study (BES), 0 represented the furthest left and 10 represented the furthest right on the spectrum. All responses were optional and a Don't Know option was also provided. A starting question also asked participants to rank their level of knowledge of British parliamentary politics on a 6-point likert scale. This question was purely exploratory and was not used to filter the responses in any way. Of the 30 MPs, 13 were members of the Conservative Party, 13 were members of the Labour Party, 2 were members of the Liberal Democrats, and 1 was a member of the Green Party. The final MP chosen was an Independent but was formerly a member of the Labour Party. There was a deliberate attempt to balance the sample of MPs as much as possible across the left/right spectrum (based on personal knowledge) as well as between more well-known MPs such as party leaders and cabinet ministers and lesser-known backbenchers.

The complete survey format distributed to participants as follows:

Q1: What would best describe your level of knowledge of British parliamentary politics?

- None
- Know a limited amount
- Know a moderate amount
- Know a fair amount
- Know a great deal
- Expert

Q2: In politics people sometimes talk of left and right. Where would you place each of these MPs on the following scale?

0 - Left

8 *et al.*

1

2

3

4

5

6

7

8

9

10 - Right

Don't Know

Q3: Where would you place each of these political parties on the following scale?

0 - Left

1

2

3

4

5

6

7

8

9

10 - Right

Don't Know

Q4: Where would you place each of these media organisations on the following scale?

0 - Left

1

2

3

4

5

6

7

8

9

10 - Right

Don't Know

Details of the 30 MPs chosen for the survey along with the results of the expert survey can be found in **Table 2** overleaf:

Table 2. Summary statistics of the expert survey ideology estimates of 30 MPs. Although the choice ranges from 0 - 10, values in the table reflect their raw value on the 11-point scale. Thus, a chosen value of 0 (furthest left) is point 1 on the scale and so on.

Name	Party	Min.	Max.	Mean	Std.D	Variance	Count
Anneliese Dodds	Labour	3	6	4.25	0.73	0.54	63
Bell Ribeiro-Addy	Labour	1	5	2.62	1.03	1.06	34
Ben Wallace	Conservative	7	10	8.38	0.70	0.49	63
Boris Johnson	Conservative	6	10	8.38	0.82	0.68	68
Caroline Lucas	Green	1	9	2.69	1.19	1.42	68
Chris Bryant	Labour	2	6	4.37	0.87	0.77	60
Desmond Swayne	Conservative	8	11	9.76	0.85	0.72	45
Diane Abbott	Labour	1	7	2.25	1.05	1.10	68
Ed Miliband	Labour	2	7	3.90	0.90	0.81	67
Edward Davey	Lib Dem	4	7	5.78	0.75	0.57	65
Hilary Benn	Labour	3	7	4.41	0.92	0.85	66
Ian Lavery	Labour	1	5	2.35	0.99	0.98	40
Jacob Rees-Mogg	Conservative	9	11	10.04	0.71	0.51	69
Jeremy Corbyn	Independent	1	4	1.67	0.76	0.57	69
John McDonnell	Labour	1	3	1.84	0.72	0.53	67
John Redwood	Conservative	7	11	10.11	0.72	0.51	63
Keir Starmer	Labour	3	8	4.81	0.89	0.79	69
Kemi Badenoch	Conservative	8	11	9.58	0.79	0.63	67
Lisa Nandy	Labour	3	7	4.17	0.81	0.66	65
Matthew Hancock	Conservative	7	10	8.06	0.81	0.65	67
Priti Patel	Conservative	8	11	9.85	0.73	0.54	68
Rishi Sunak	Conservative	7	11	8.58	0.89	0.79	69
Steven Baker	Conservative	4	11	9.63	1.03	1.06	63
Stuart Andrew	Conservative	4	11	8.00	1.65	2.73	11
Suella Braverman	Conservative	9	11	10.22	0.68	0.47	68
Theresa May	Conservative	7	10	8.07	0.71	0.51	68
Thomas Tugendhat	Conservative	7	10	7.82	0.81	0.66	62
Tim Farron	Lib Dem	4	8	5.79	1.02	1.03	67
Yvette Cooper	Labour	3	7	4.62	0.79	0.63	66
Zarah Sultana	Labour	1	5	2.07	0.98	0.95	45

Additionally, details of the 12 political parties and 13 media organisations can also be found in **Table 3:**

Table 3. Summary statistics of the expert survey ideology estimates of 12 political parties and 13 media organisations. Although the choice ranges from 0 - 10, values in the table reflect their raw value on the 11-point scale. Thus, a chosen value of 0 (furthest left) is point 1 on the scale and so on.

Name	Group	Min.	Max.	Mean	Std.D	Variance	Count
Conservative	Party	7	11	8.71	0.71	0.50	68
Labour	Party	2	7	4.50	0.78	0.60	68
SNP	Party	2	8	4.29	0.97	0.94	68
Liberal Democrats	Party	4	7	5.74	0.74	0.55	68
Green	Party	1	6	2.96	0.97	0.94	67
Reform	Party	5	11	10.05	0.94	0.89	55
Plaid Cymru	Party	2	8	3.95	1.10	1.20	64
Alba	Party	2	11	5.24	2.38	5.65	38
SDLP	Party	1	6	4.31	0.92	0.84	48
DUP	Party	6	11	9.67	1.05	1.09	55
Sinn Fein	Party	1	5	2.96	0.86	0.74	51
Alliance	Party	3	6	5.38	0.70	0.48	48
BBC	Media	4	9	5.98	0.80	0.64	64
ITV	Media	4	8	6.17	0.73	0.54	60
Sky	Media	4	9	6.59	0.88	0.77	61
The Daily Express	Media	7	11	9.74	0.88	0.78	65
The Daily Mail	Media	8	11	9.91	0.72	0.51	65
The Daily Mirror	Media	1	9	3.88	1.16	1.34	65
The Daily Telegraph	Media	7	11	9.53	0.84	0.70	66
The Financial Times	Media	4	10	6.71	1.19	1.41	63
The Guardian	Media	2	6	3.86	0.80	0.63	66
The Huffington Post	Media	2	8	4.59	1.13	1.27	41
The Independent	Media	3	7	4.96	1.08	1.16	57
The Sun	Media	6	11	9.02	0.91	0.83	64
The Times	Media	6	11	8.02	1.07	1.14	64

5. Sept 2022 Conservative Leadership Contest - Model Variables

To formally model the ideological component to leadership endorsements, a simple binary logistic regression was used where the model predicts support for eventual winner, Truss, relative to the unsuccessful candidate, Sunak. A recent paper by Jeffery et al. (2023) examined the candidate preferences of Conservative MPs in the September 2022 contest, exploring a wide range of personal, political and ideological factors that may have influenced the vote. Their study found that, contrary to common media narrative, loyalty to previous leader Boris Johnson was *not* a significant factor in determining support for Truss. However, they do note the importance of ideological factors in this contest. In place of the ideological proxies they use in their model such as ERG membership and support for the levelling-up agenda, this model used an MP's ideal point. Beyond that, the also controlled for a simple set of demographic and political variables, in-keeping with relevant factors also present in previous studies by Jeffery et al. (2018) and Jeffery et al. (2022) on the voting motivations of Conservative MPs in the 2016 and 2019 party leadership contests.

Model Variables

The four demographic variables chosen to control for were an MP's age, gender, ethnicity, and educational background. Social variables are primarily sourced from the *Representative Audit of Britain* (RAD) - a dataset collating together research on parliamentary candidates in Great Britain - drawing on demographic data for MPs elected as part of the 2019 parliamentary cohort.¹ This includes their gender, ethnicity, year of birth, their secondary school type [State, Grammar, Private], and their university type [None, Non-Russell Group, Russell Group, Oxbridge]. For a handful of MPs elected after 2019 via by-elections, these data are manually sourced from publicly available online sources. Three variables relating to an MP's political background are selected for inclusion in the model: election cohort, ministerial background, and size of constituency majority. Election cohort data can also be derived from the RAD dataset and is categorised into five distinct groups [Pre-1997, 1997-2010, 2010-2015, 2015-2019, 2019]. Ministerial background is sourced manually from publicly available online sources and categorised into three groups: [Never (been a minister), Former, Current (at time of the contest)]. Constituency majority size is also sourced manually from publicly available online sources and is treated as a continuous variable (%). Summary statistics of these variables for MPs who endorsed each candidate are shown in **Table 4**.

Table 4. % is calculated as a percentage of complete observations for each variable by candidate.

	Sunak		Truss	
	N	%	N	%
Vote Share	130	-	150	-
Social Variables				
Gender				
Male	106	82%	106	71%
Female	24	18%	44	29%
Ethnic Minority				
No	124	95%	141	94%
Yes	6	5%	9	6%
Year of Birth (Median)	52.1	-	51.5	-
School Type				
Private	56	49%	51	41%
Grammar	19	17%	19	15%
State	39	34%	53	43%
University Type				
Oxbridge	40	33%	28	21%
Russell Group	42	35%	46	35%
Non-Russell Group	33	27%	41	31%
None	6	5%	17	13%
Political Variables				
Cohort				
Pre-1997	4	3%	6	4%
1997 - 2010	29	22%	21	14%
2010 - 2015	37	28%	37	25%
2015 - 2019	28	22%	40	27%
2019	32	25%	46	31%
Ministerial Position				
Current	29	22%	46	31%
Former	41	32%	33	22%
Never	60	46%	71	47%
Majority (%) (Median)	28.8	-	27.9	-

6. Ordinary User Ideal Points

This appendix consists of details about the ideal points generated by the CA follower model for the row data (ordinary users). When applying correspondence analysis to the MP follower adjacency matrix, the CA model will scale both the columns *and* the rows meaning that ideal points are also generated for the ordinary users who follow the MPs. **Figure 5** illustrates the ideal point distribution of these users (who will follow 10 or more MPs):

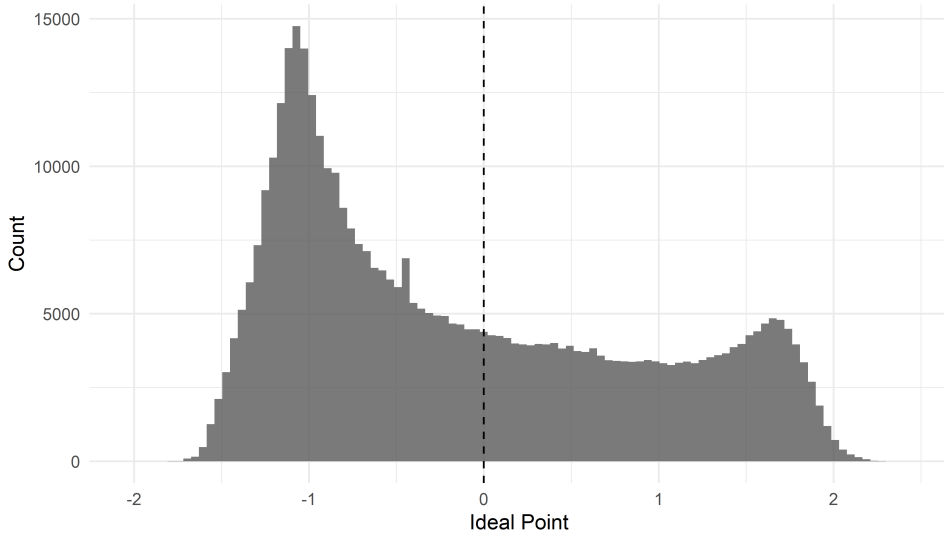


Figure 5. Histogram distribution of ordinary user ideal points (who follow at least 10 MPs). Bins = 100.

Along the distribution, it is interesting to note an imbalance between users to the left compared to the right. Given that the liberal skew of Twitter is well-documented, this is perhaps an intuitive finding and likely also explains the imbalance between the number of followers of MPs from left-wing parties compared to right within the dataset. Although these users represent an especially informative subset of profiles (follows at least 10 MPs), they can still be regarded as ordinary users. To test the general ideological distribution of wider elite users on the platform (e.g: journalists, media organisations, political commentators etc.), this set of followers is subsetting to only include 'elite' accounts. This is done by filtering the follower profiles to only include accounts which are either verified (before the creation of *Twitter (X) Blue*, where users can now pay a monthly subscription to receive verified status, verification was an official signal of authenticity) and/or have at least 30,000 followers (This is in accordance with ruling by the Advertising Standards Authority (ASA) - <https://www.asa.org.uk/rulings/sanofi-uk-A19-557609.html>). This included 11,525 users and **Figure 6** illustrates their ideal point distribution:

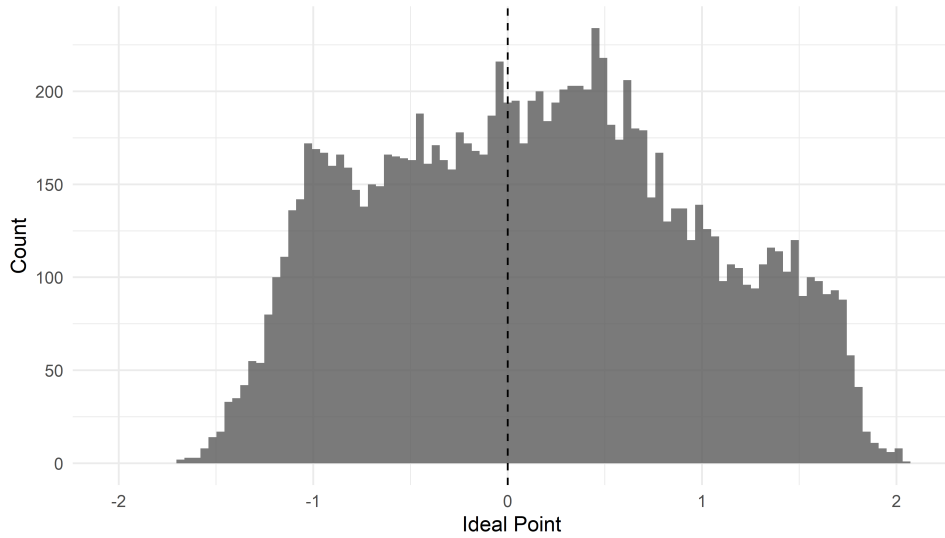


Figure 6. Histogram distribution of wider elite user ideal points (verified and/or have a min. 30,000 followers). Bins = 100

This subset of elite users demonstrate a more balanced distribution of ideal points across the spectrum, with a slight remaining over-representation of left-wing elite users on the site.

References

- Barberá, P. 2015. Birds of the same feather tweet together: bayesian ideal point estimation using twitter data. *Political Analysis* 23 (1): 76–91. <https://doi.org/10.1093/pan/mpu011>.
- Barberá, P., J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau. 2015. Tweeting from left to right: is online political communication more than an echo chamber?. *Psychological Science* 26 (10): 1531–1542. <https://doi.org/10.1177/0956797615594620>.
- Greenacre, M. J. 2010. Correspondence analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (5): 613–619. <https://doi.org/10.1002/wics.114>.
- Jeffery, D., T. Heppell, R. Hayton, and A. Crines. 2018. The conservative party leadership election of 2016: an analysis of the voting motivations of conservative parliamentarians. *Parliamentary Affairs* 71 (2): 263–282. <https://doi.org/10.1093/pa/gsx027>.
- Jeffery, D., T. Heppell, and A. Roe-Crines. 2022. The conservative party leadership election of 2019: an analysis of the voting motivations of conservative parliamentarians. *Parliamentary Affairs* 75 (1): 113–134. <https://doi.org/10.1093/pa/gsaa046>.
- Jeffery, D., T. Heppell, A. Roe-Crines, and C. Butler. 2023. Trusting truss: conservative mps' voting preferences in the (first) british conservative party leadership election of 2022. *Representation*, <https://doi.org/10.1080/00344893.2023.2231469>.