

Research Design

Different data sources data set were used:

- **European Social Survey** [ess2016]. The ESS is a multi-country scientific survey conducted every two years since 2002 by scientists in several European countries. The objectives of the ESS are to understand the changing attitudes and values in Europe, to explain how European institutions are changing and to develop a range of European social indicators and values [ess08]. The data for the present analysis were collected from different survey waves the latest for each country.
- **Chapel Hill Expert Survey** [ches2014]: Since 1999 CHES provides party positioning scores on European integration, ideology and policy issues for national parties in a variety of European countries. The more recent survey waves also comprise questions on non-EU policy issues, such as immigration, redistribution, decentralization, and environmental policy.
- **Varieties of Democracy** [vdem2017]: V-DEM aims to transparently produce Indicators of Democracy that are multidimensional and disaggregated. This modern framework reflects the complexity of the concept of democracy and captures seven high-level principles of democracy: electoral, liberal, participatory, deliberative, egalitarian, majoritarian and consensual, and collects data to measure these principles.

The dependent variable from our final model is the populism cluster inferred from Ches party data that has been matched and merged to the ESS micro data on voting a particular party. Variable Description ...

- summary stats

Model Based Clustering

Voting for a right wing party is often approximated by left-right scales or is time-intensive coded. This analysis adopts Cas Mudde’s clear minimalist definition of populism to identify core features that all sub types of populism have in common. In line with this definition, we suggest that populist parties are primarily shaped by their degree of anti-establishment attitudes as well as their opposition to globalization. Subsequently, we propose to classify European populist parties along a progressive and traditionalist left-right dimension. Some CHES party indicators are part of the clustering even though they are not present in the ESS micro data.

This multidimensional classification problem is best approached by model-based hierarchical clustering [mclust]. This tool set is already adopted in political science and praised for estimating meaningful clusters on high-dimensional data political science [mclust_pol1; mclust_pol2]. Model-based clustering assumes the data generating process to be driven by a mixture of underlying probability distributions in which each component represents a different cluster. Consequently each sub population is separately estimated and summarized by a mixture of these sub populations. Therefore the density tends to be centered at the multidimensional means (μ_k) and increased by geometric features (shape, volume, orientation) of the clusters being determined by the parameters of the covariance matrices Σ_k , which may also induce cross-cluster conditions [mclust_pol1]. In a nutshell the Gaussian Finite Normal Mixture model assumes a d-dimensional data set y_1, \dots, y_n to calculate G components with the likelihood

$$\ell_{MIX}(\theta_1, \dots, \theta_G | y) = \prod_{i=1}^n \sum_{k=1}^G \tau_k f_k(y_i | \theta_k)$$

where f_k is the density of the θ_k mixture parameters, unlike traditional methods model-based clustering uses a soft assignment and calculates τ_k that represents the probability of a given observation belonging to the k component [mclust_pol1]. Unsupervised machine learning algorithms are often criticized for introducing bias by hyper parameter settings. mclust addresses this by estimating a grid of different models and hyper parameter constellations. Bayesian Information Criterion (BIC) is provided to pick the most useful model [mclust]. The metric is penalized for the complexity of the model to ensure Occam’s Razor.

The estimated grid models are represented by the following identifiers: EII, VII, EEI, VEI, EVI, VVI, EEE, EEV, VEV and VVV. The first Letter refers to volume, the second to shape and the third to orientation. E stands for “equal”, V for “variable” and I for “coordinate axes” [kassambara2017practical].

There is a clear indication of a four-component mixture with covariances having different shapes but the same volume and orientation (EVE) (VEV (ellipsoidal, equal shape)).

Following Milligan and Cooper’s (1988) finding that standardization by dividing each variable by its range gives consistently superior recovery of the underlying cluster structure, all the variables are standardized by dividing by each variable’s range. PCA ...

Table X displays the characteristics of each cluster through mean and standard deviation.

To validate the clusters let’s inspect the party classification for three countries. For checking the consistency of the clusters, the variable means are calculated by cluster in Table @ref(tab:cluster_means). The final cluster vector is attached to the data with name **cluster**.

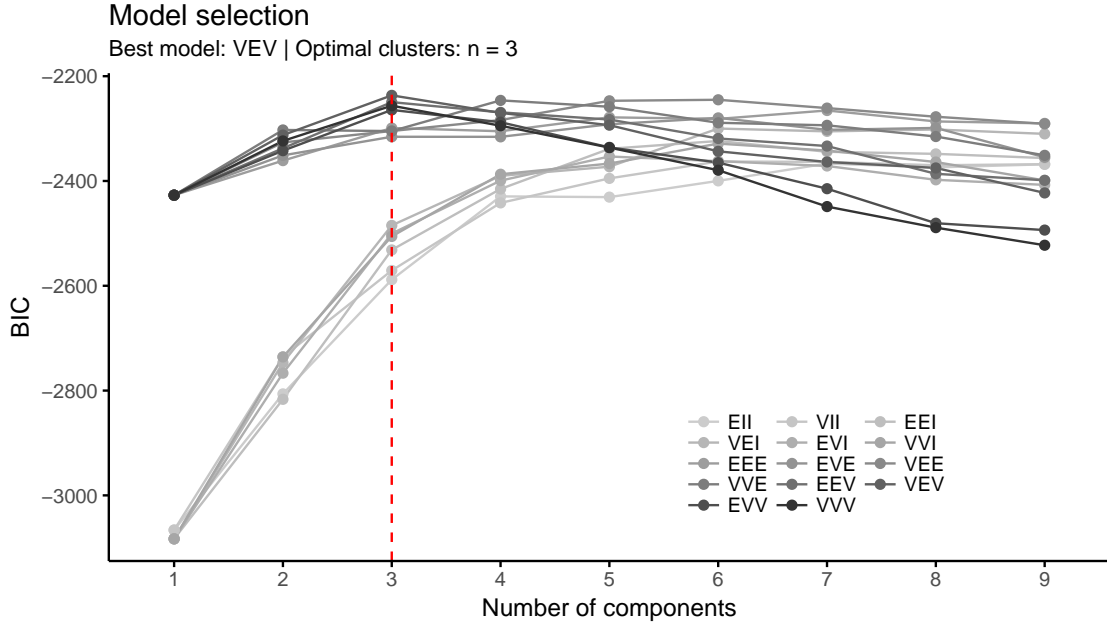


Figure 1: BIC Model Selection for mixture models on the CHES party positioning data set. Following Fraley and Raftery (2010), the grid component structures are labeled using: EII=spherical, equal volume; EEI=diagonal, equal volume and shape; EVI=diagonal, equal volume, varying shape; EEE=ellipsoidal, equal volume, shape, and orientation; VEV=ellipsoidal, equal shape; VII=spherical, unequal volume; VEI=diagonal, varying volume, equal shape; VVI=diagonal, varying volume and shape; EEV=ellipsoidal, equal volume and equal shape; VVV=ellipsoidal, varying volume, shape, and orientation.

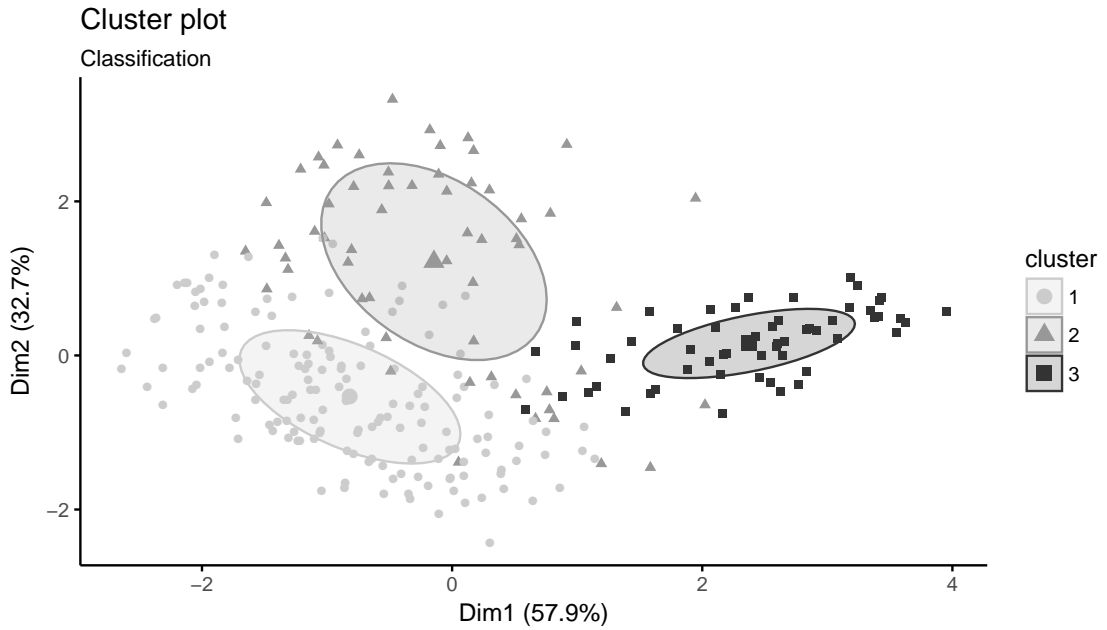


Figure 2: Classification and Cluster Boundaries

Table 1: Cluster means and standard deviations for populist indicators

Var	Establishment	Left Populist	Right Populist
antielite_salience	3.42 (1.78)	5.9 (2.44)	7.73 (1.5)
civlib_laworder	4.67 (1.75)	3.47 (1.97)	8.19 (1.07)
eu_position	6.16 (0.51)	3.28 (1.15)	2.81 (1.24)
galtan	4.52 (2.13)	3.58 (2.36)	8.35 (0.95)

¹ Note: Standard deviations are in parentheses

Table 2: Party Cluster Membership

Cluster	Parties
Establishment	CDU (ger); SPD (ger); FDP (ger); Grunen (ger); CSU (ger); PS (fra); PRG (fra); EELV (fra); UMP (fra); MODEM (fra); NC (fra); PRV (fra); AC (fra)
Left Populist	Linke (ger); Piraten (ger); DieTier (ger); PCF (fra); PG (fra); Ens (fra)
Right Populist	NPD (ger); AfD (ger); FN (fra); MPF (fra)

¹ Note: ...

Principal Component Analysis

Principal Component Analysis (PCA) another unsupervised machine learning approach is now used to reduce the dimensions of two item batteries to get a parsimonious model size. For both variables we are only interested in the first Dimension/Component that contributes most to the variability in the underlining data structure. The formal model for the first principal component of a data set is the linear combination its features

$$Z_i = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}X_p$$

that has the largest variance and where is the first principal component loading vector, with elements $\phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}$ are normalized, which means that $\sum_{j=1}^p \phi_{j1}^2 = 1$. To calculate these loadings, we must find the vector that maximizes the variance. It can be shown using techniques from linear algebra that the eigenvector corresponding to the largest eigenvalue of the covariance matrix is the set of loadings that explains the greatest proportion of the variability. This method does not depend compared to Factor Analysis or SEM on model assumptions or multivariate error distributions. Despite all dimensionality reduction methods have different background and purposes they often yield similar results. But for construct validation I recommend factor analysis or any other framework that provides hypothesis testing.

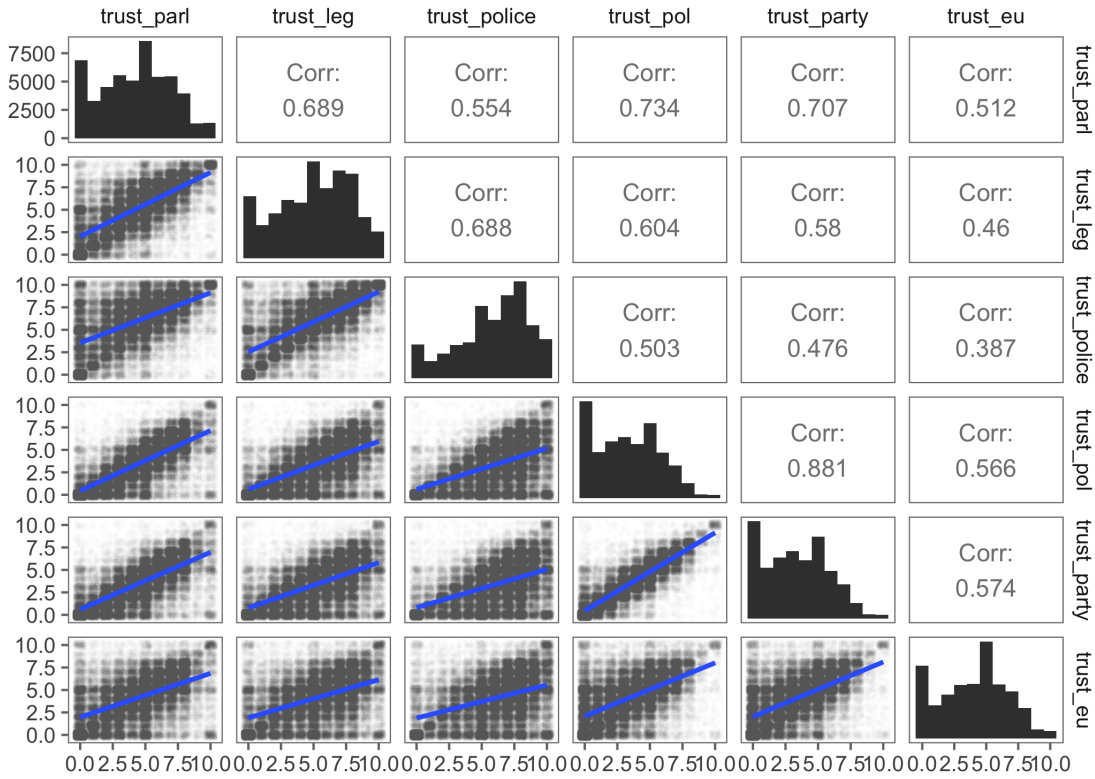


Figure 3: Bivariate Scatter plots for item battery public trust

Trust Items

First we explore the diemnionality of the given trust items by applying standard pairwise scatter plots for each variable combination. The visual inspection confirms a strong linear association between all variables

and the diagonal histograms approximate a normal distribution with a heavy tail on 0, as people are especially critical against political institutions and put more overall confidence in the police.

The next plot shows a two-dimensional representation of the data that captures most of the information in a lower-dimensional subspace. The First two components contribute together to nearly 80% of the total variance.

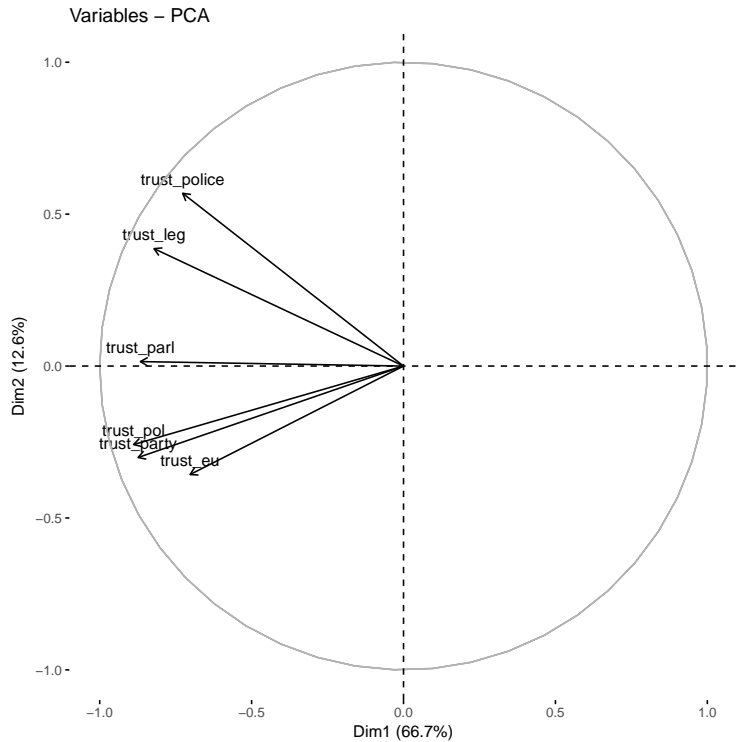


Figure 4: First and Second PC Dimensions for the Trust Items

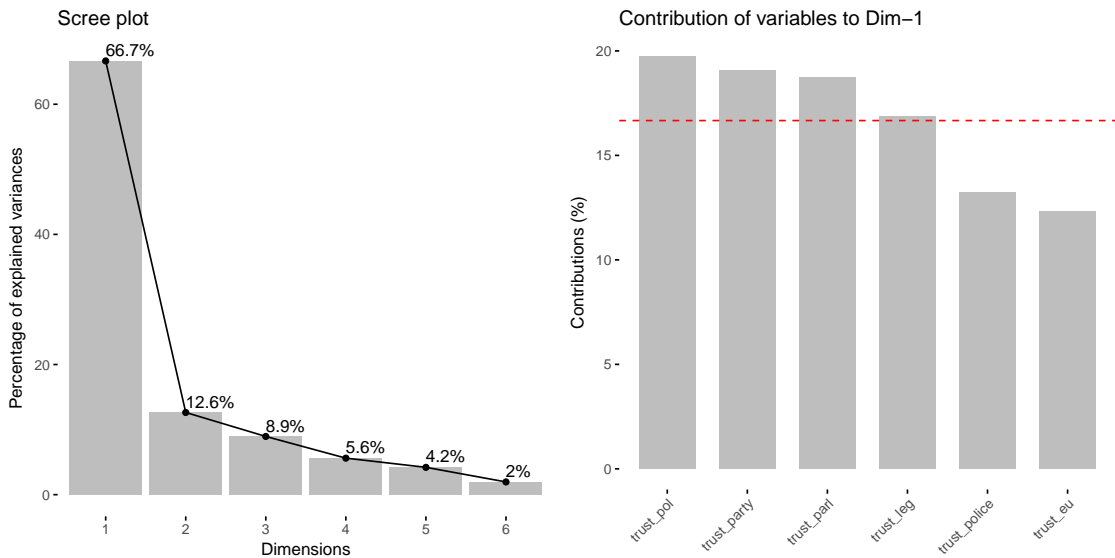


Figure 5: scree plot and Component Contribution by Variable

As each principal component vector defines a direction in the feature space and all arrows in Figure X point to the same direction we can be confident to approximate one trust dimension. This notion is supported by the scree plot capturing 66,7% of the total Variance by the first dimension. We can further see that the trust items for political institutions do somewhat differ in their contribution to the first PC due to different reasons. The European Parliament is often seen as disconnected from people and the police is always highly appreciated by everyone. The first PC scores are stored for later analysis as `pc_trust`.

Immigration Items

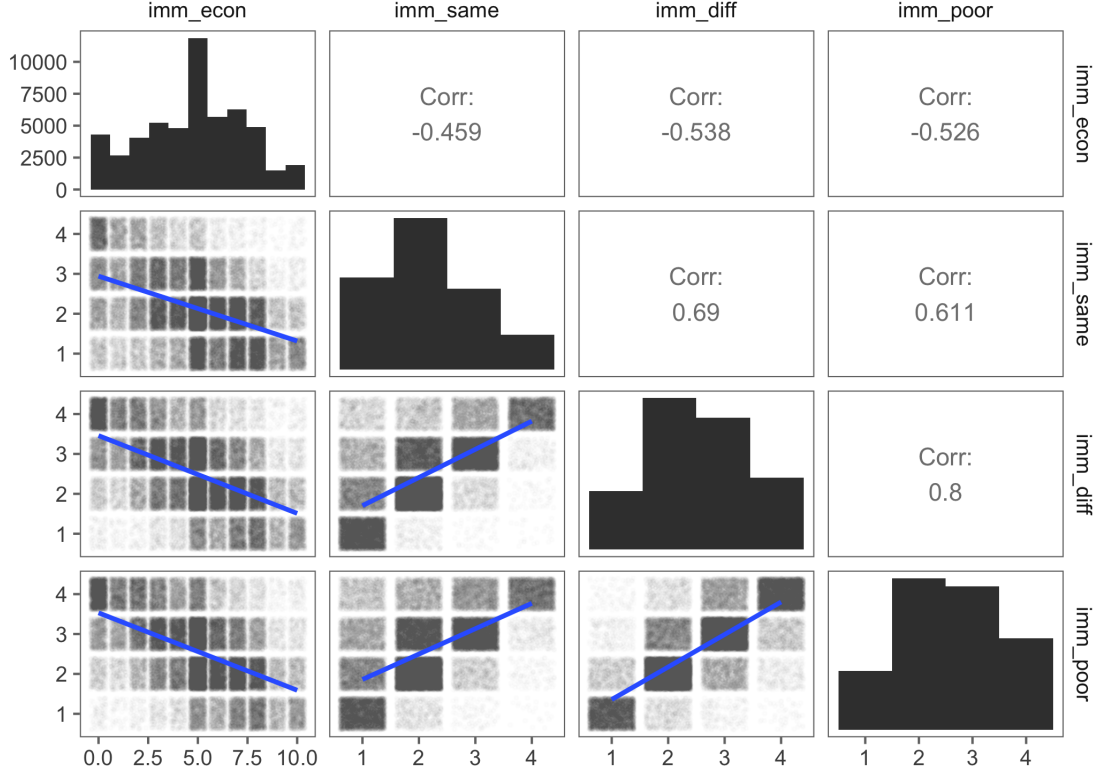


Figure 6: Bivariate Scatter Plots for item battery immigration

Next we applied the same PCA for the Immigration Items. But first we explore again the nationality of the given data by standard pairwise scatter plots for each variable combination in figure @ref(fig:imm_cor). The visual inspection again confirms a strong linear association between all variables but seem to be more noisy than the trust items before. The diagonal histograms approximate a normal distribution, despite the fact that three out of four variables have only 4 levels. This Likert scale is assumed to be equal-distant to be suitable for PCA.

The two-dimensional representation of the immigration items proposes again a strong one PC solution but points to a minor second dimension entirely rooted in `imm_econ`. We are only interested in the first PC that captures over 70% of the total variance. The scree plot and contribution plot support these findings. The final PC scores are stored for later analysis as `pc_imm`.

Statistical Models

As detailed explained in section X and the dependent variable y_i is labeled as 1 for voters of right wing populist parties estimated by model based-clustering.

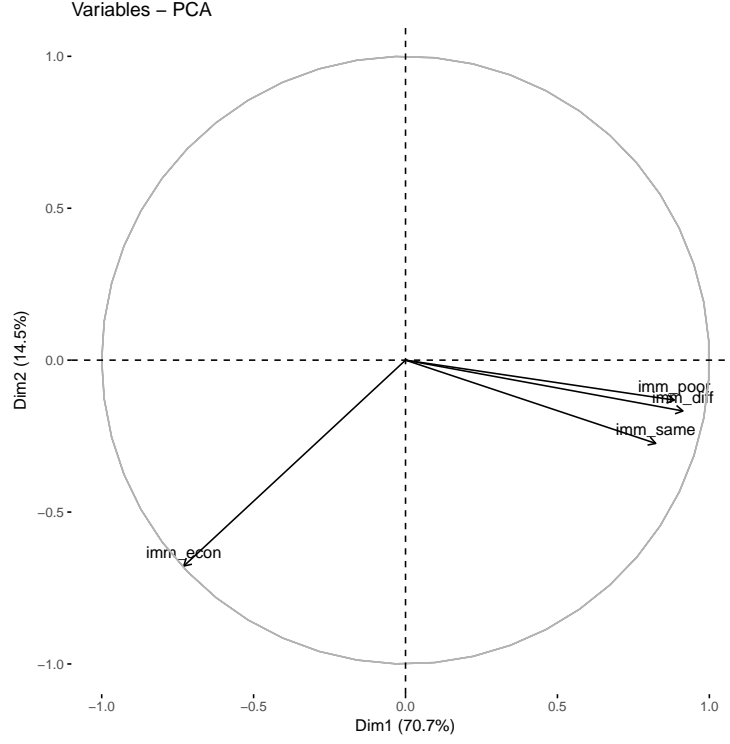


Figure 7: PCA Dimensionality for the Immigration Items

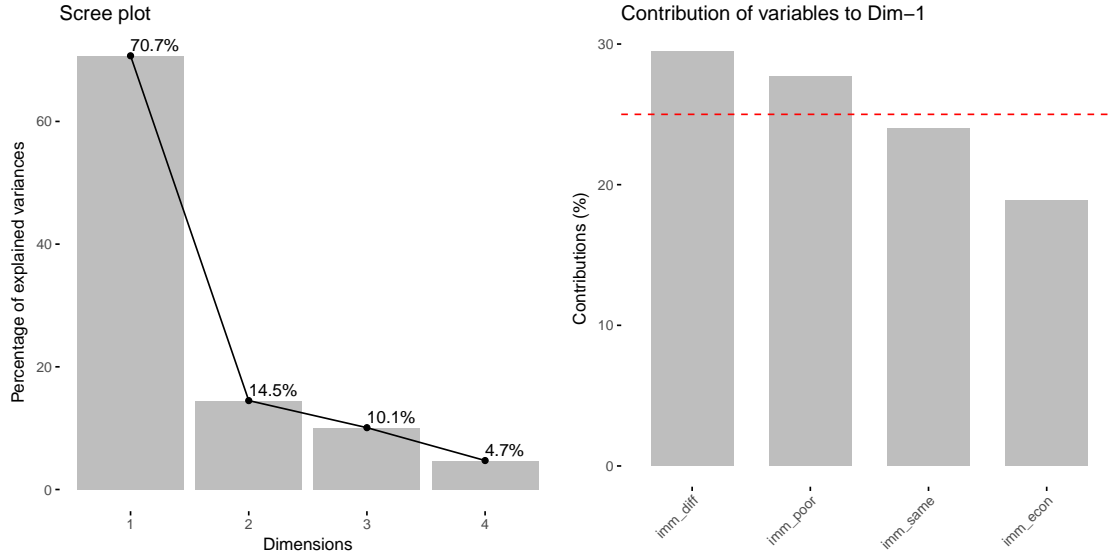


Figure 8: scree plot and Component Contribution by Variable

$$y_i = \begin{cases} 1 & \text{if voted for a right populist party} \\ 0 & \text{if not} \end{cases}$$

This binary data structure is best fitted by generalized linear models (GLMs) with Binomial response function (Bernoulli). As part of this paper we built different logistic regression models coming from different statistical backgrounds. First we used the common frequentist approach to statistical inference but rapidly augmented

Bayesian models brought by the `rstanarm` package. Second we employed multilevel models both frequentist and bayesian to better capture the nested data generation process of the ESS micro/country data structure [Gelman07].

Frequentist Approach

The first simple logistic model (frequentist) predicts the mean log-odds of a respondent voting right-populist, through a linear combination of an intercept and a slope the latter allowing to quantify the effect of each covariate.

$$y_i \sim \text{Binomial}(n_i, p_i) \quad (1)$$

$$\text{logit}^{-1}(p_i) = \alpha + \beta x_i \quad (2)$$

The inverse link function can be switched to the right-hand side of the formula to non-linearly transform the outcome.

This complete pooling architecture is separately deployed to each country dataset and summarized accordingly to explore country-specific model deficiencies. The fitting function is taken from the core `stats` function of R.

The simple version of the GLM assumes independence of each individual observation which produces in a less severe scenario downwards biased parameter standard errors (SE). By deploying random effects with context level predictors we are able to model country-specific heterogeneity and group level standard errors [Steenbergen02]. The formal multilevel logistic regression has the formula

$$y_i \sim \text{Binomial}(n_i, p_i) \quad (3)$$

$$\text{logit}^{-1}(p_i) = \alpha_j + \beta x_i \in (0, 1) \quad (4)$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \quad (5)$$

The group intercepts α_j are assumed to be an independent and identically distributed random variable with hyperprior grand mean and equal variance. This population distribution is partially fitted by pooling its sub-population intercepts (share information) and used to refit or regularize the parameters towards 0. Thereby we compromise between complete pooling and no pooling that in order to balance bias and variance (trade-off). In sum we avoid fixed effects for nested data to control model complexity and prevent overfitting.

The models discussed so far are often seen in social sciences. Now we leave the world of asymptotic significance and turn to the Bayesian framework.

Bayesian Approach

These models have substantive advantages for our analysis as

- estimates are penalized towards a plausible parameter space
- prior knowledge can be incorporated
- do not link evidence to p-values but rather propagate the uncertainty of the model to the posterior distribution so we can inspect it.
- ... and many more.

Normal Bayesian logistic regression

$$y_i \sim \text{Binomial}(n_i, p_i) \quad (6)$$

$$\text{logit}^{-1}(p_i) = \alpha + \beta x_i \in (0, 1) \quad (7)$$

$$\alpha \sim N(0, 10) \quad (8)$$

$$\beta \sim N(0, 10) \quad (9)$$

Varying Intercept ...

$$y_i \sim \text{Binomial}(n_i, p_i) \quad (10)$$

$$\text{logit}^{-1}(p_i) = \alpha_j + \beta x_i \in (0, 1) \quad (11)$$

$$\alpha_j \sim N(\alpha, \sigma) \quad (12)$$

$$\alpha \sim N(0, 10) \quad (13)$$

$$\sigma \sim \text{HalfCauchy}(0, 10) \quad (14)$$

$$\beta \sim N(0, 10) \quad (15)$$