Measuring Political Polarization in Mass Publics

The Cluster-Polarization Coefficient

Isaac D. Mehlhaff (mehlhaff@live.unc.edu)

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Objective: Measure the polarization of a distribution consistent with the two dynamics invoked in the political science literature: distance and concentration of component groups.

Cluster-Polarization Coefficient

For each individual i in cluster k who holds a position on dimension j:

$$CPC = 1 - \frac{\sum_{k=1}^{n_k} \sum_{i \in C_k} \sum_{j=1}^{n_j} (x_{i_{kj}} - \mu_{k_j})^2}{\sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (x_{i_j} - \mu_j)^2} \cdot \frac{1}{n_k}$$

$$\forall \quad i \in (1, ..., n_i), \quad j \in (1, ..., n_j), \quad k \in (1, ..., n_k)$$

Analytical Design

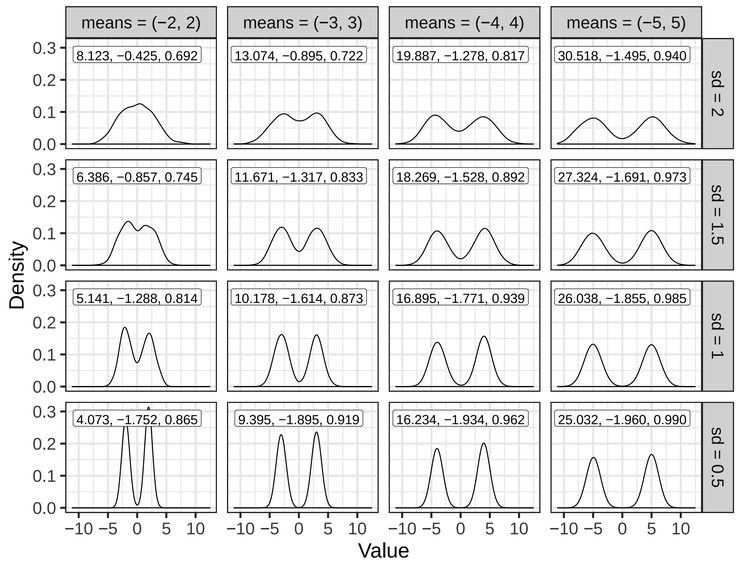
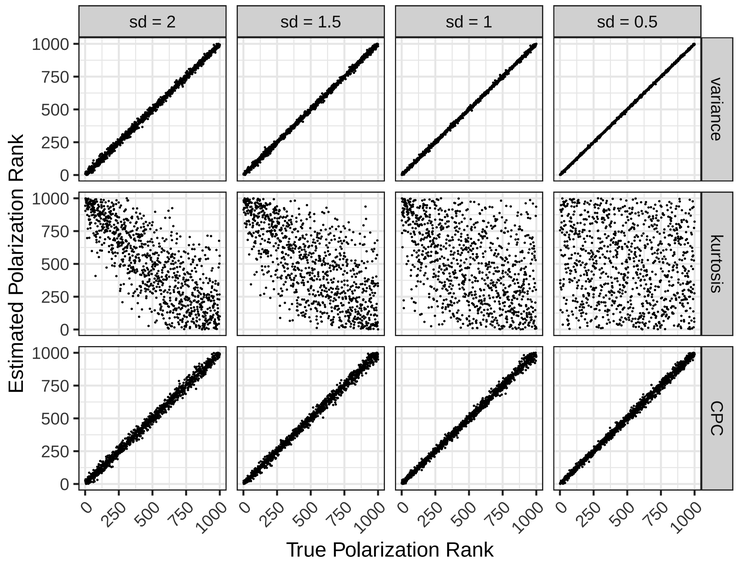


Figure 1: Polarization simulated by randomly manipulating means (distance dynamic) and standard deviations (concentration dynamic) of Gaussian mixtures. Labels on each facet report global variance, global kurtosis, and CPC, respectively.

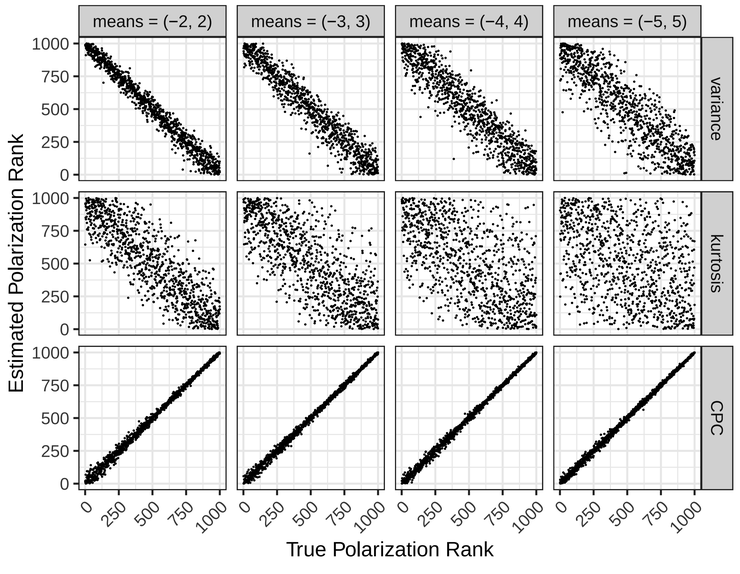
Simulation Results

		distance	concentration	overall
univariate	variance	15.469	572.354	404.863
	kurtosis	575.495	576.254	575.875
	CPC	33.797	29.603	31.769
bivariate	variance	10.983	568.973	402.4
	kurtosis	504.499	525.794	515.256
	CPC	21.654	17.744	19.795

Table 1: Root mean squared error for Monte Carlo simulations, broken down by dynamic.



(a) Random means



(b) Random standard deviations

Figure 2: Bivariate Monte Carlo simulations of polarization measures, comparing true rank order of distributions with predicted rank order based on polarization measures; $y = x$ line represents perfect performance.

Substantive Results

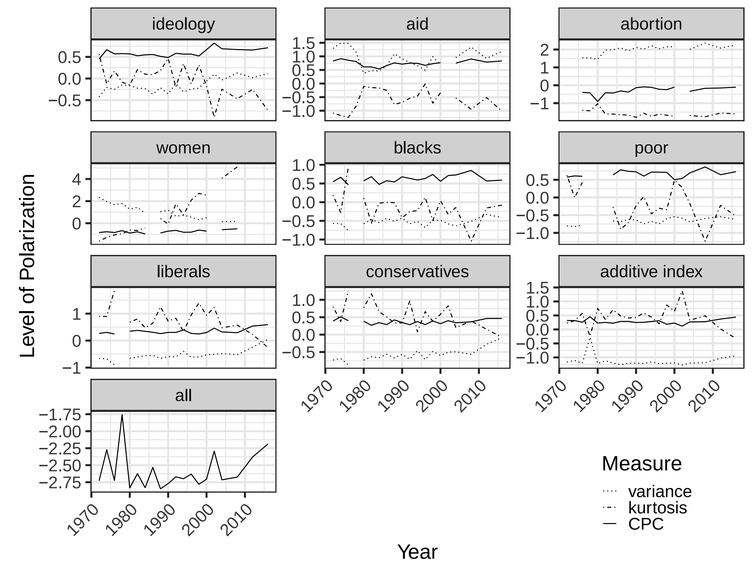


Figure 3: ANES polarization estimates over time for ideology, policy positions, and feeling thermometers (from DiMaggio, Evans, and Bryson (1996)). Measures scaled to $\sim N(0, 1)$ to enable comparison across variables, across measures, and over time. Facet "all" displays polarization estimates based on all variables in a seven-dimensional issue space.

Highlights

- The CPC possesses two desirable properties: It is naturally bounded on the domain $[0, 1]$ and it comports with theoretical dynamics of polarization. That is, it increases when the distance between clusters increases, when clusters become more tightly concentrated around their centroid, or when the size of polarized groups increases.
- The CPC performs as expected on both dynamics of polarization, outperforms variance and kurtosis — the two most popular measures in the literature — and improves slightly in higher dimensions.
- Because the CPC can be scaled to a high-dimensional space, it can take into account numerous variables without needing to create additive indices (which discard variation in the data, regress to the mean, and lack identifiability) or use item-response models.
- The choice of measure matters; especially after the early 2000s, the CPC disputes conclusions DiMaggio, Evans, and Bryson (1996) would have made had the data been available to them.