

Political Polarization via Bayesian Updating

Emmet Hall-Hoffarth and Afonso Rodrigues

November 1, 2021

1 Model

At $t = 0$ a set C of n agents start with and $(h \times 1)$ vector of prior beliefs over a set of h issues which together form the $(n \times h)$ matrix π_0 . At each time step $t > 1$ every agent will simultaneously update their beliefs according to the following procedure. Consider, without loss of generality, the case of agent 0. At the beginning of the period agent 0 will select an "in group"; a subset of other agents $I \in C$ of size $|I|$ who the agent will move their beliefs towards. Other agents are selected into c without replacement with a probability inversely proportional to their distance in views to agent 0 (in l2-norm) and proportional to the l2-norm of that agent's belief vector, as in equation (1):

$$P_0(i \in I) \propto \frac{\|\pi_i\|_2^\epsilon}{\|\pi_i - \pi_0\|_\delta} \quad (1)$$

Where ϵ is an "extreme bias" parameter, and δ is a "endogenous selection" parameter. This model embodies two mechanisms. Firstly, agents are more likely to choose other agents with similar views in their consideration set. This probability updates at every time t , so it captures the idea that agents may endogenously choose the views that they are exposed to, for example, on social media. Secondly, agents are more likely to consider the views of agents who lie further from the centre. We call this "extreme bias." Intuitively, this reflects the incentive of media organizations to report on controversial topics or views because they generate more viewer engagement. The relative strength of these effects depends on the parameter ϵ . The parameter δ captures how agents decide who to pick in their consideration set. If $\delta \rightarrow \infty$ then the agents select uniformly without prejudice from all agents. If $\delta \rightarrow 0$ then agents give lexicographic preference to agents with whom they have the exact same view on the largest number of dimensions (although this will never happen in practice, as views are continuous). In general, we parameterize $\delta \in (0, 1)$, to capture the following intuition: we believe that agents are more likely to consider the views of other agents particularly when they are close on most issues, even if they are far away on a small number of them.

Furthermore, each period every agent will select an "out group" $O \in C$ whose views the agent will move away from. The out group is selected in the

opposite manner to the in group; agents are more likely to be selected if they are further away in $l\delta$ distance:

$$P_0(i \in O) \propto \|\pi_i\|_2^\epsilon \cdot \|\pi_i - \pi_0\|_\delta \quad (2)$$

Having chosen a consideration set, agent 0 now applies a simple updating rule: they update their belief in period $t + 1$ to be the geometric mean of their prior beliefs and the mean of their consideration set. The equation for this is shown in (3):

$$\pi_{0,t+1} = \pi_{0,t} + \frac{\overline{\pi_{I,t}} - \pi_{0,t}}{4} - \alpha \frac{\overline{\pi_{O,t}} - \pi_{0,t}}{4} \quad (3)$$

where $\alpha > 0$ is a parameter that regulates the relative strength of the in and out group effects.

2 Possible Extensions

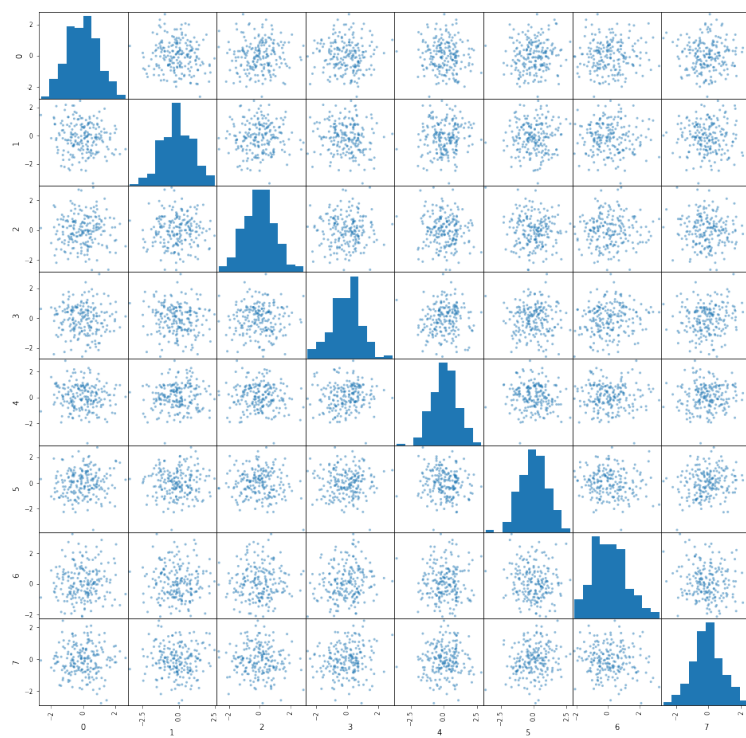
- Not all agents update at every step (stochastic updating).
- Some proportion of agents are re-randomized every period (i.e. due to birth and death) (already implemented in code)
- Allow agents to have heterogeneous selection and updating rules.
- Consider the effect of high dimensions (h) and extreme bias (ϵ) on the observed amount of polarization
- Find a metric to quantify polarization.

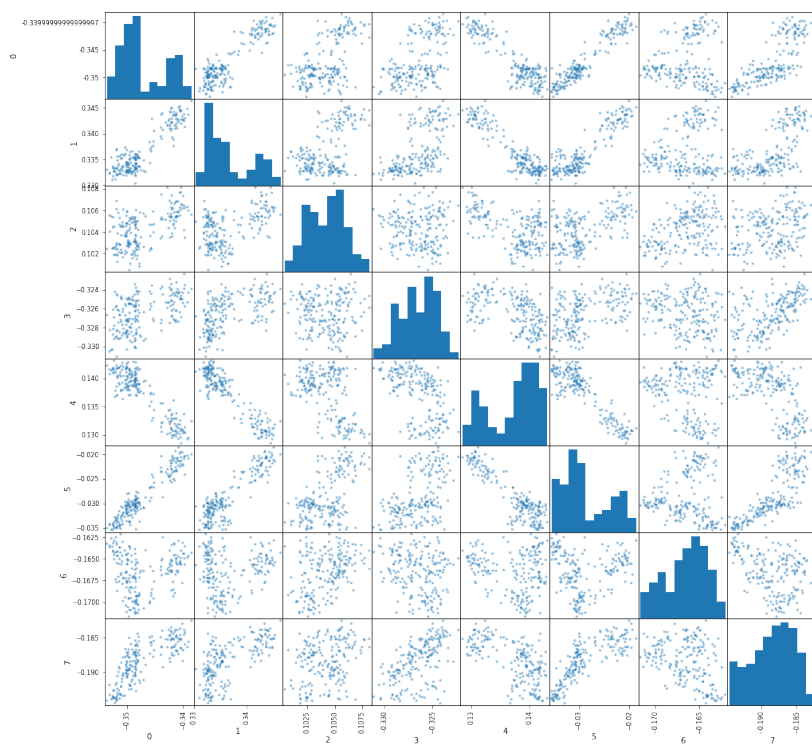
3 Results

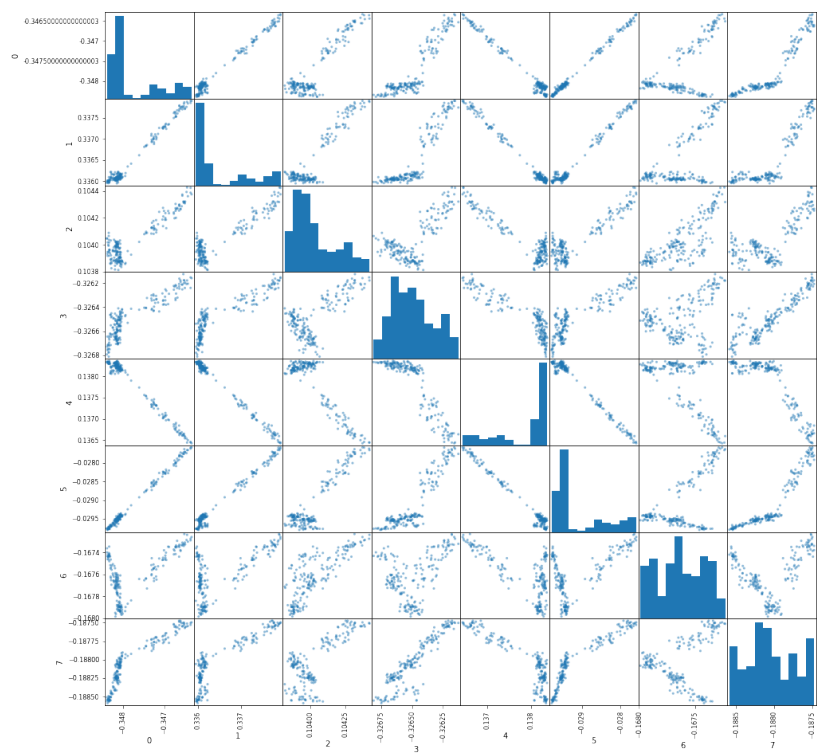
We simulate the model over 50 iterations with the following parameters:

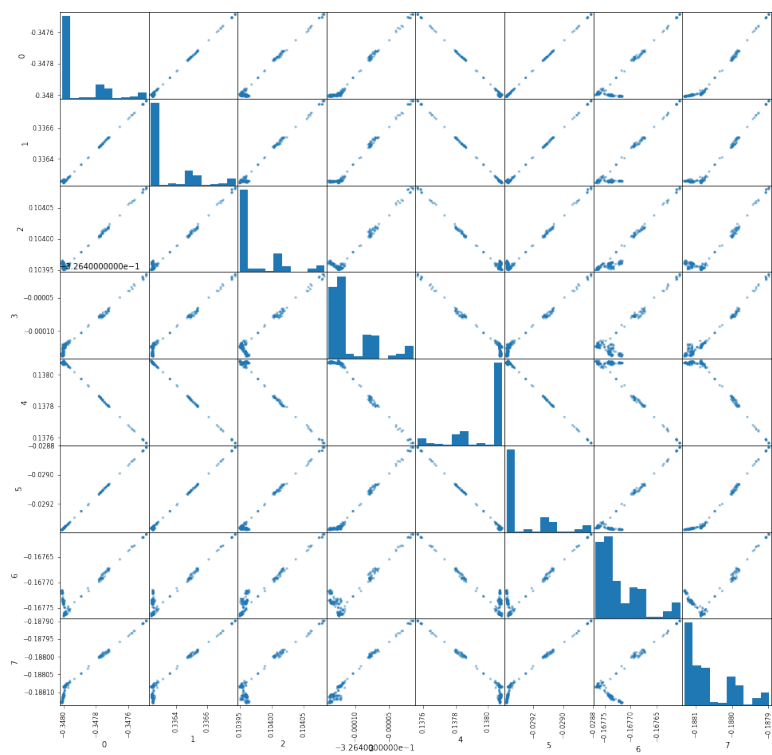
- $n = 500$
- $|I| = |O| = 10$
- $h = 2$
- $\epsilon = 1.5$

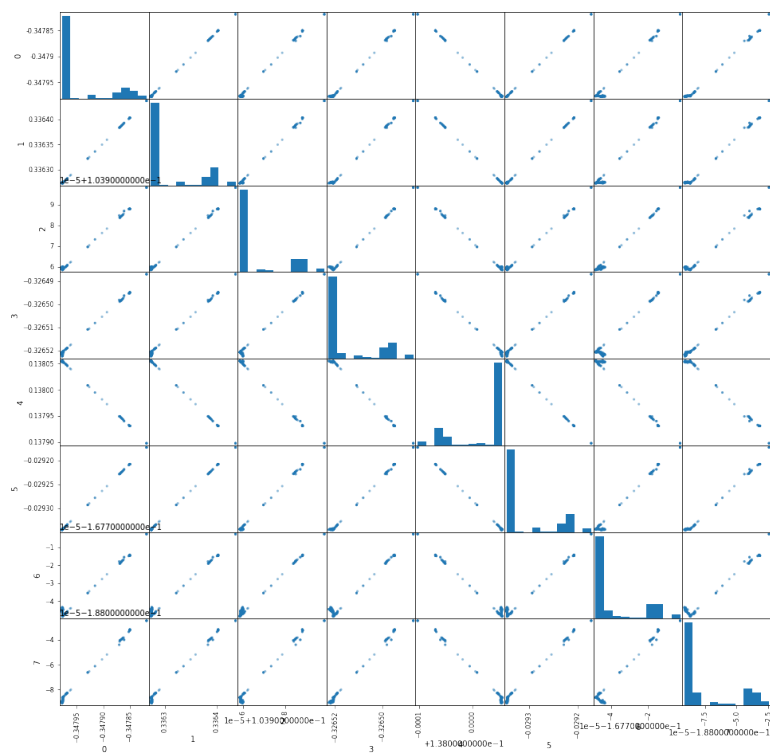
We obtain the following results over 50 iterations (each image is shown 5 iterations after the last, with the exception of the last image, which is 100 iterations after the former):

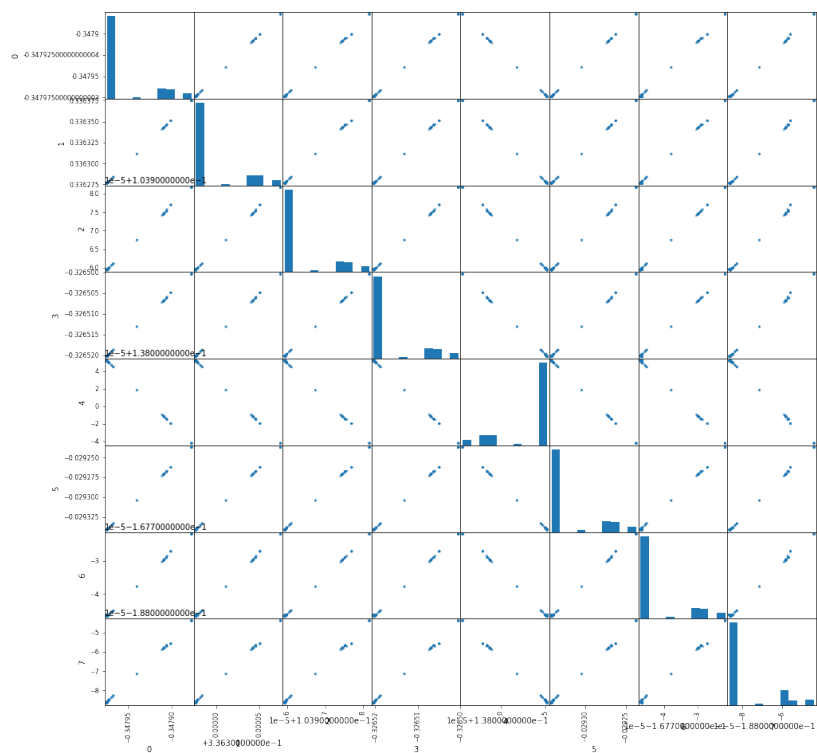


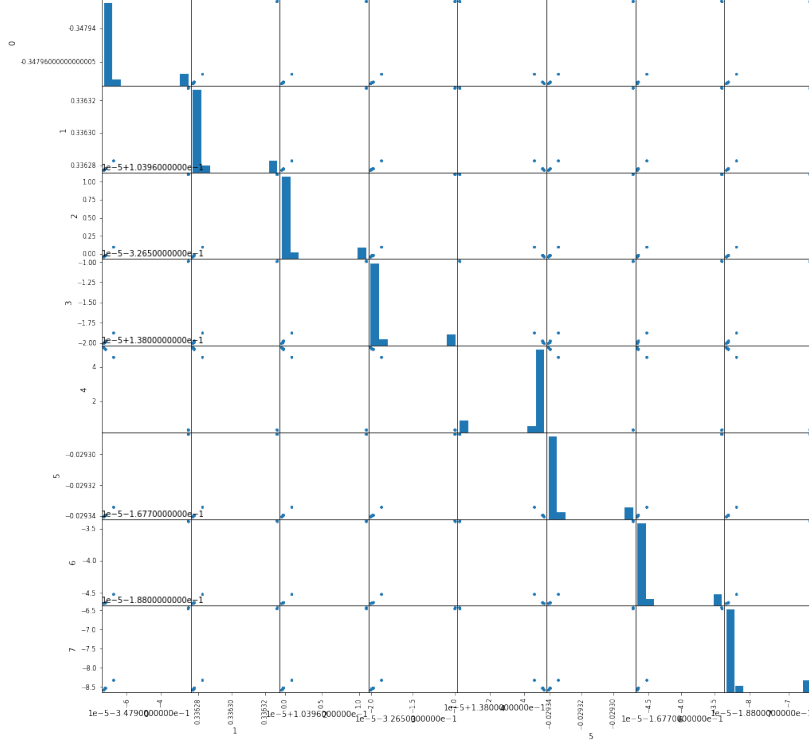












We can observe in these results two distinct forms of polarization. The first is a sort of dimensionality reduction (perhaps call this multidimensional polarization). The initially uncorrelated space of views eventually becomes essentially one-dimensional as agents line up along the diagonal in each cross-section. Initially, there are perhaps some kinks or triangular shapes, but eventually they fall back down into the dominant line. In this model there is a strong tendency towards conformity. The other is "absolute polarization," whereby we see agents moving out towards extreme positions. Especially after a large number of iterations, we see that the model converges to a state where agents hold essentially only one of two views. The relative strength of these effects will depend on the parameterization of the model, exactly how remains to be investigated.