Examining the Viability of Markov Chains to Model Political Terror Scale Score Distributions

Devanshi Shah, First-year Undergraduate Political Science Student

Introduction

Created in 1976, the Political Terror Scale (PTS) is a system used to measure the level of political repression and conflict a country experiences in a year based on the dimensions of scope, intensity, and range. The primary sources of data upon which the scores are coded are annual reports published by Amnesty International, Human Rights Watch, and the U.S. State Department. If a report by one of these entities is published within a given year, then it will be coded into an individual PTS score on a scale of one to five, with five being the most egregious repression. For example, if each reporting agency publishes its findings on a given country, then for that year, the PTS will display three separate scores based on the content of each report; scores are not averaged into a single number so as to account for the inevitable discrepancies in reporting.

The PTS contains data on all 195 countries in some capacity (scoring gaps exist for certain years), which means that there has been a nearly consistent flow of PTS data for most countries. Markov models are a useful method for extracting meaningful patterns from these scores, especially when trying to discern trends at the macro-level over a considerable period of time. A Markov chain can help us determine potential future evolution of the distribution of PTS scores and how their convergence to some type of equilibrium would appear.

Markov chains are stochastic models that use probability theory to predict a sequence of events over a time series, in which probabilities are constructed based on the state an agent attains in the previous time-step. The essential components of a Markov chain are the states, which are possible vectors the system contains; the set of states makes up the state space S. Transition probabilities predict the probability of some agent transitioning from State i to State j. They exist for every possible combination of state transitions from time step N to N+1, which means that the sum of the probability values for each state must sum to 1.0, or 100%. A matrix can be employed to model the full extent of transition probabilities, with a value assigned to each possible combination of time step N to N+1. Assuming N represents the column states, and N+1 represents the row states, each column should sum to 1.0.

Markov chains do not take into account the history of the system, and instead only rely on agents' immediately previous states to predict their movement for the following time step; this process is iterated over a time series. Thus, it is important to be cautious when using a Markov chain to model social science systems such as the PTS score patterns, because the history of the system does not inform projected movement. Eventually, over a time series that is long enough (depending on the system), the system will reach some equilibrium point where the distribution of agents in each state remains constant or nearly constant after every time step. This is called the stationary distribution. In order for the stationary distribution to be attained, according to the Perron-Frobenius Theorem, 4 conditions must be met:

Assumption 1:
There is a finite set of states.

Assumption 2:
Transition probabilities are fixed.

Assumption 3:

The system is ergodic- agents can get from any one state to another through a sequence of transitions.

Assumption 4:
The system is not a simple cycle.

The prediction of PTS score distributions may fit well into a Markov chain, notably when considering how the system reaches a stationary distribution.

Central Question: *How does a Markov chain capture the overall dynamics of score shifting and distributions from 1976 to 2021 for countries entered into the PTS system?*

Sub-Questions:

1. *Are countries with specific scores more or less prone to receiving consistent scores over time? What can explain these trends?*
2. *Are there any significant differences in the stationary/equilibrium distributions for each of the three datasets, and what accounts for these differences?*
3. *Can this model be used as a sound predictor of future score distributions?*
4. *Are there any interventions that could change the course of the stationary distributions permanently?*

Methodology

Before calculating the matrices, it is necessary to define the major parameters of the Markov chain. The states are simply each possible PTS score, so there are five states ranging from one to five. The transition probabilities will represent the likelihood of a country moving from one state to another in the system. Ergo, it will represent the probability of a state's PTS score changing on a year-to-year basis.

To dissect the differences in reporting and scoring between each of the three reporting agencies (Amnesty International, Human Rights Watch, and the State Department), the optimal strategy would be to create transition matrices for each dataset. Then, they can be averaged to obtain some sort of final transition matrix that accounts for the overall macro trends, after understanding the individual results of the respective matrices. To accomplish this, the PTS data was cleaned and formatted into three separate tables containing yearly scores: one for each dataset.

After obtaining each dataset, the transition probabilities for each matrix were calculated using the existing patterns in the data tables themselves. Through the for-loop function in Python, the transition probabilities for each possible combination of scores were calculated from all the years available from each dataset by dividing a specific score combination of two consecutive years (i.e. one to one, one to two, one to three, one to four, one to five) for each state and dividing by the total number of the original score (state) for the first year (i.e. the number of ones assigned in the first year). Then, each of the obtained probabilities for all the two consecutive year pairs were placed into their own matrices (i.e. for Amnesty International, there were nearly 45 matrices because the data ranged from 1976 to 2021). Then, these matrices were averaged to obtain the final matrices for each of the three sources.

After obtaining four usable matrices in total, two methods were employed to find the stationary or equilibrium distributions that each matrix would produce. The first directly found the stationary distribution π through linear algebra, specifically eigenvalue decomposition to produce π as a five by one matrix in which each vector adds up to 1.0 ($\pi = [\ \pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5] = 1.0$). The second method was to calculate the distribution iteratively using the previously referenced for-loop and the initial distribution for each dataset (marginal probabilities for the first year in each table) and running the Markov chain given the transition matrix to see how the distribution finds an equilibrium over several iterations. Both methods yielded the same results when done for each of the four matrices, confirming the accuracy of the stationary distribution calculations.

Results

The matrices are as follows:

Amnesty International:

```
array([[0.7778339 , 0.20158078, 0.01750441, 0.00308091, 0.        ],
       [0.11622957, 0.65340683, 0.21208463, 0.01738908, 0.00088989],
       [0.01299412, 0.21587532, 0.63416587, 0.12765282, 0.00931187],
       [0.00066138, 0.02082644, 0.22643035, 0.60856428, 0.14351755],
       [0.        , 0.0038966 , 0.02379357, 0.25856717, 0.71374266]])
```

Human Rights Watch:

```
array([[0.63249008, 0.32361111, 0.01041667, 0.03348214, 0.        ],
       [0.11117638, 0.64014241, 0.24086872, 0.0078125 , 0.        ],
       [0.        , 0.12724146, 0.70875237, 0.15685747, 0.00714869],
       [0.        , 0.00543478, 0.22341701, 0.66929671, 0.10185149],
       [0.        , 0.        , 0.03461538, 0.1395542 , 0.82583042]])
```

State Department:

```
array([[0.84847155, 0.13211055, 0.01894348, 0.00047441, 0.        ],
       [0.1687095 , 0.65912078, 0.16508226, 0.00635797, 0.00072949],
       [0.00488798, 0.24241893, 0.63633111, 0.10967476, 0.00668722],
       [0.00409854, 0.00844534, 0.3924829 , 0.41524515, 0.17972807],
       [0.        , 0.        , 0.04056647, 0.27050721, 0.68892632]])
```

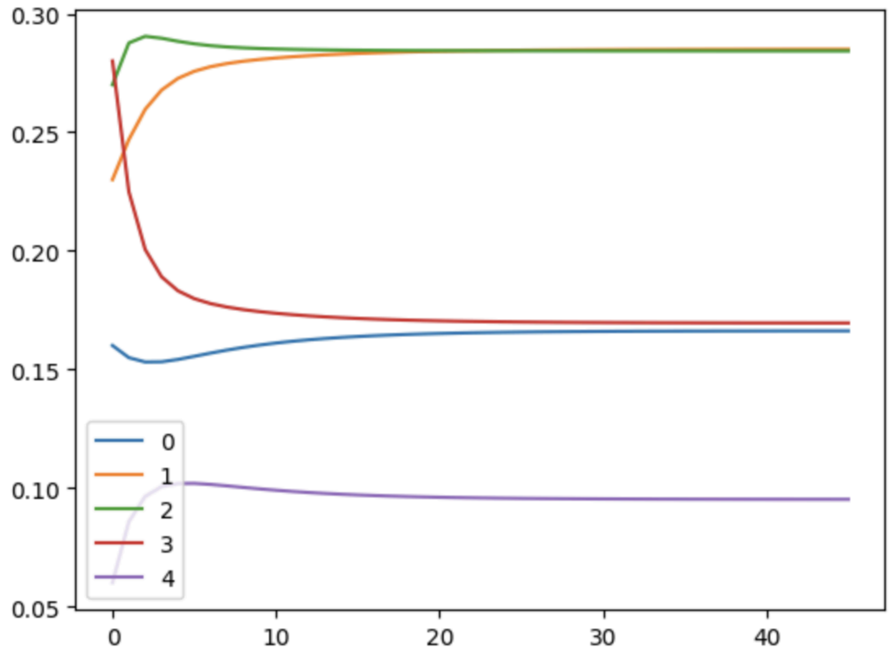
Then, these three matrices were averaged to obtain the “final” matrix:

```
array([[0.75293184, 0.13203848, 0.0059607 , 0.00158664, 0.        ],
       [0.21910081, 0.65089001, 0.19517857, 0.01156886, 0.00129887],
       [0.01562152, 0.20601187, 0.65974978, 0.28077675, 0.03299181],
       [0.01234582, 0.01051985, 0.13139502, 0.56436872, 0.22287619],
       [0.        , 0.00053979, 0.00771593, 0.14169904, 0.74283313]])
```

The calculated stationary distributions are as follows:

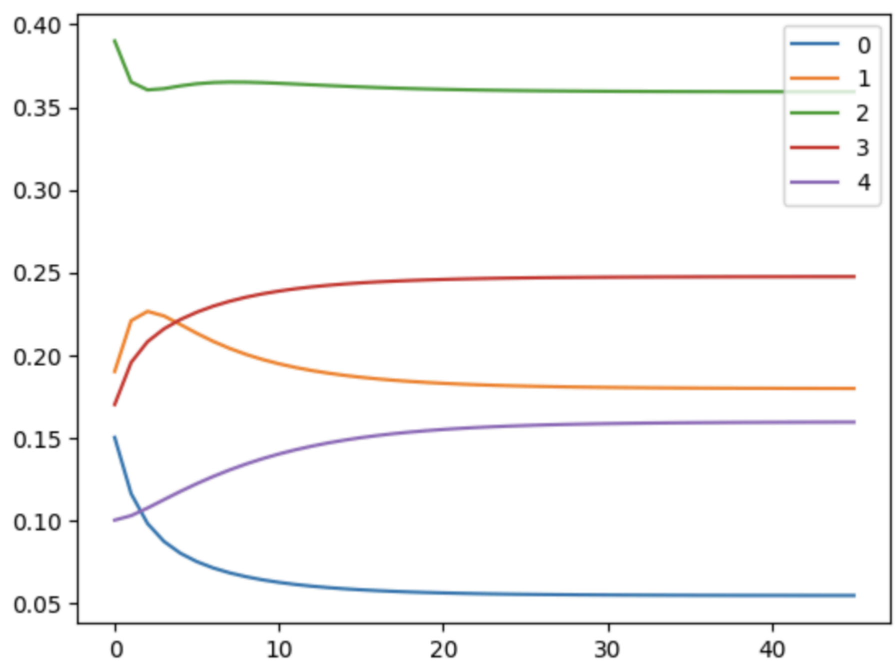
Amnesty International:

```
array([0.16621093, 0.28496106, 0.28423949, 0.16948412, 0.0951044 ])
```



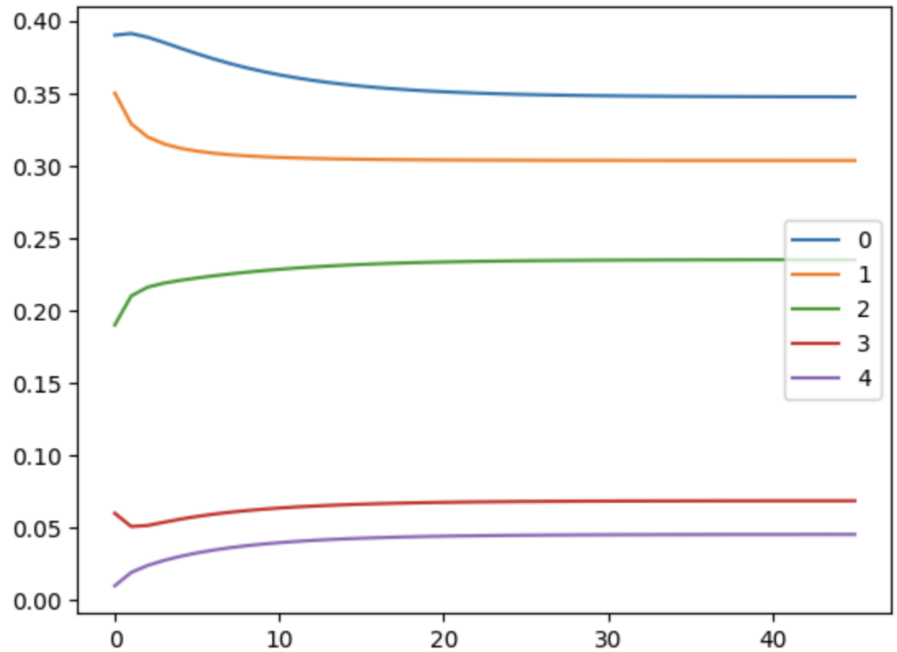
Human Rights Watch:

```
array([0.05433624, 0.17961645, 0.35922214, 0.24740358, 0.1594216 ])
```



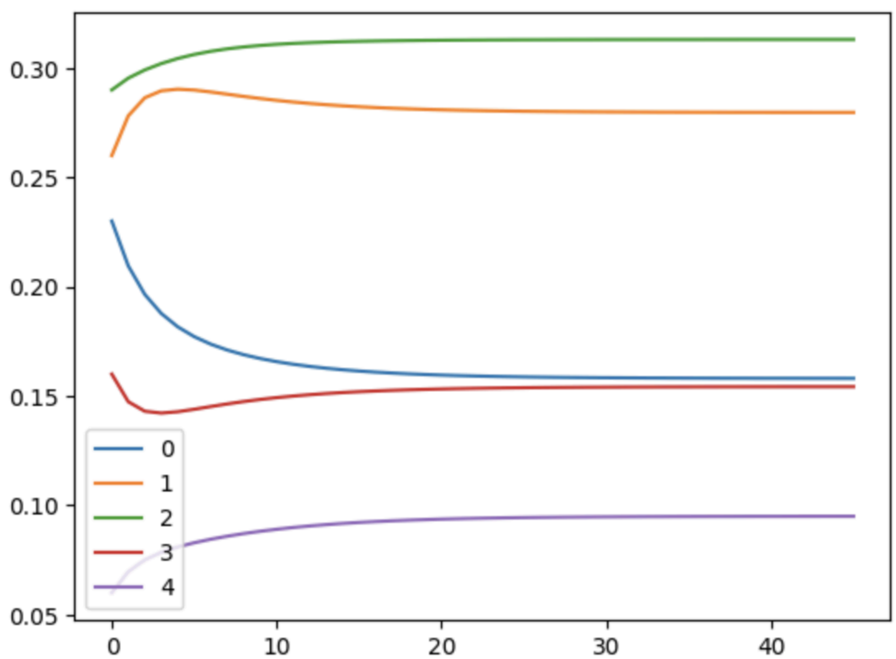
State Department:

```
array([0.3473215 , 0.30347185, 0.23505696, 0.06869516, 0.04545452])
```



“Final” Stationary Distribution (averaged):

```
array([0.15800579, 0.27967007, 0.3130901 , 0.15425734, 0.0949767 ])
```



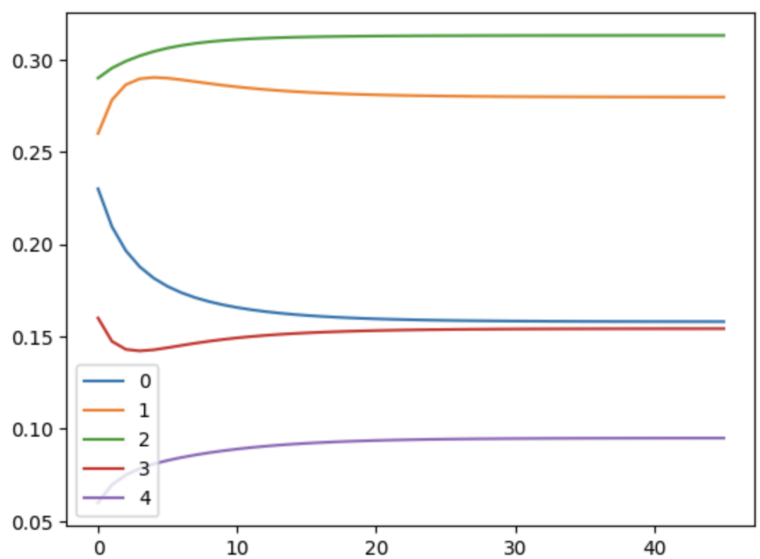
Analysis (with elements of REDCAPE)

Using the guiding questions from the Introduction to assess the outputs of the model, there are several interesting patterns that arise both within and among the matrices and stationary distributions.

Central Question: *How does a Markov chain capture the overall dynamics of score shifting from 1976 to 2021 for countries entered into the PTS system?*

Analyzing the final matrix obtained in the previous section (pictured below for convenience), the obvious conclusion here is that the majority countries in any given state are most likely to remain in the same state the following year as evidenced by the diagonal transition probabilities being the greatest in magnitude of each column. If a country does shift states, then it is most likely to shift states to a score that is within a margin of one of the original score. Overall, the farther a score is in magnitude from the original state/score, the less likely a state is to shift to that 'farther" score at the next interval. This makes sense, because it is reasonable for states to stay relatively consistent in terms of political terror; drastic changes from year to year, such as going from a 5 to 1, should be rare considering the massive level of intervention needed to significantly reduce repression. If a state does see a drastic change in PTS scores, this change likely occurred over a long period of time, where transitions between states occurred with the common "flow" of countries from one state to another. The results of the matrix are what was expected prior to constructing the model.

```
array([[0.75293184, 0.13203848, 0.0059607 , 0.00158664, 0.        ],
       [0.21910081, 0.65089001, 0.19517857, 0.01156886, 0.00129887],
       [0.01562152, 0.20601187, 0.65974978, 0.28077675, 0.03299181],
       [0.01234582, 0.01051985, 0.13139502, 0.56436872, 0.22287619],
       [0.        , 0.00053979, 0.00771593, 0.14169904, 0.74283313]])
```



Examining the stationary distribution of the final matrix (pictured above), the graph demonstrates that after around 20 years (given the initial distribution), the proportions of countries in each state converge to a constant value from year to year. What is interesting is that the distribution of 1s (least repressive) noticeably decreases over time, while the distribution of 5s converges after growing during the initial iterations. In contrast, the distributions of other states seem to have been more stable from the start. If anywhere from a 2 to a 3 is considered an "average" score for a country, then the stationary distribution here supports this conclusion. At any given time period according to the stationary distribution, a country is most likely to be in the "state" of score 3, followed by 2. It is least likely to be a 5, which makes sense, given that in a world where democratic norms have been taking root, only a small proportion of countries would be classified as fives.

Another interesting trend is the magnitude of transition probabilities within the earlier described diagonal. It is established that countries are most likely to retain their previous year's score for the next year, but what is even more notable is that countries within either a one or a five (the two extremes of the spectrum) are the most likely to remain in the same state than any of the other states (two, three, or four). If a country's score is either a 1 or 5, the probability it retains that score is about 75 percent, whereas this probability for the other states are around 65, 66, and 56, respectively. One possible explanation can be that countries that receive a score of one are typically also liberal democracies (i.e. Sweden), which tend to be stable and secure over time in their governance, thus rarely being engulfed in political terror. On the other hand, countries that receive a five (i.e. Syria) are typically authoritarian in nature, and these nations ironically tend to maintain a consistent level of instability and repression of the most egregious forms. It would be very rare that a repressive authoritarian regime would roll back political terror, and even if improvements were made, the score would likely stay within a magnitude of one from the original state.

Meanwhile, countries in any of the states in between these extremes seem to be more prone to fluctuating to one score up or down, and this may be due to the fact that these countries tend to be illiberal or partial democracies in practice, which means governance is more prone to be subject to inconsistencies because of the combined elements of democracy and authoritarian principles (a great example of this is India). Thus, while the levels of repression overall remain mostly constant, minute fluctuations are more common because of what can often be "messy" governing and management.

Sub-Questions:

Are countries with specific scores more or less prone to receiving consistent scores over time? What can explain these trends?

This question was just answered above, where it was determined that countries with the two most extreme scores are more prone to receive consistent scoring over time, but let's observe the breakdown for each individual dataset.

Pictured below are the matrices for Amnesty International, HRW, and the State Department (in order, with rows summing to 1.0):

Amnesty International:

```
array([[0.7778339 , 0.20158078, 0.01750441, 0.00308091, 0.        ],
       [0.11622957, 0.65340683, 0.21208463, 0.01738908, 0.00088989],
       [0.01299412, 0.21587532, 0.63416587, 0.12765282, 0.00931187],
       [0.00066138, 0.02082644, 0.22643035, 0.60856428, 0.14351755],
       [0.        , 0.0038966 , 0.02379357, 0.25856717, 0.71374266]])
```

Human Rights Watch:

```
array([[0.63249008, 0.32361111, 0.01041667, 0.03348214, 0.        ],
       [0.11117638, 0.64014241, 0.24086872, 0.0078125 , 0.        ],
       [0.        , 0.12724146, 0.70875237, 0.15685747, 0.00714869],
       [0.        , 0.00543478, 0.22341701, 0.66929671, 0.10185149],
       [0.        , 0.        , 0.03461538, 0.1395542 , 0.82583042]])
```

State Department:

```
array([[0.84847155, 0.13211055, 0.01894348, 0.00047441, 0.        ],
       [0.1687095 , 0.65912078, 0.16508226, 0.00635797, 0.00072949],
       [0.00488798, 0.24241893, 0.63633111, 0.10967476, 0.00668722],
       [0.00409854, 0.00844534, 0.3924829 , 0.41524515, 0.17972807],
       [0.        , 0.        , 0.04056647, 0.27050721, 0.68892632]])
```

The matrix for Amnesty International is healthily consistent with the overall trend, where the probability of remaining in the extreme state is at least 70 percent for both scores. For Human Rights Watch, the probability of remaining a five is extremely high, while the probability of remaining a one is consistent with the same pattern for the rest of the scores. The opposite is true for the State Department's matrix. This means that states that are relatively terror-free are more likely to fluctuate between one and two if the scores are coded based on HRW data, and states that endure the most political terror are more likely to fluctuate between a 4 and 5 if the scores are coded according to State Department data.

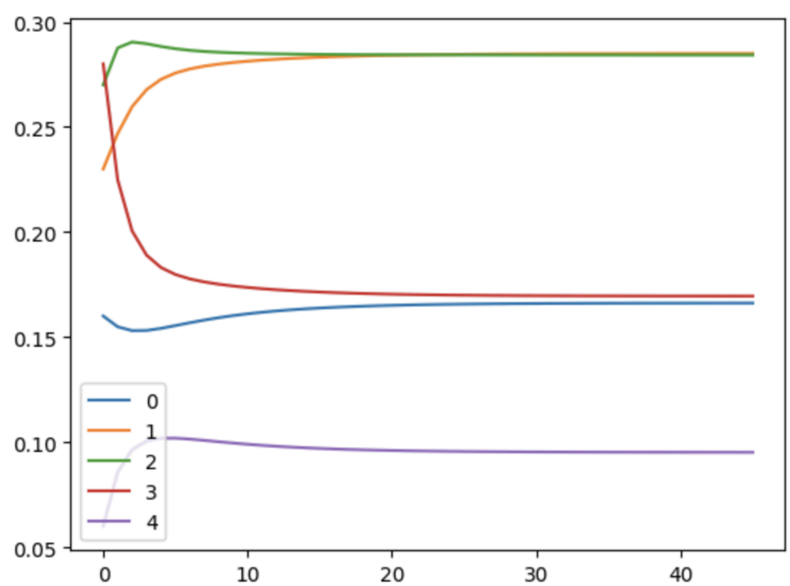
A possible explanation for this lies in the nature of reporting by each entity- perhaps, the data collection from year to year for more repressive states by the State Department is less consistent, and the same may be true for the least repressive states for HRW's data collection (and vice versa for the second extreme score). This suggests that the State Department may conduct more lax reporting that gives states the benefit of the doubt, whereas HRW is more stringent with its standards and reporting, and is more "tough" on countries where political terror is rampant. Moreover, State Department data may not be the most reliable because the United States has an interest in maintaining political alliances with states (i.e. Saudi Arabia, Israel, etc.) and downplaying serious repression in these countries to advance its geopolitical interests.

Are there any significant differences in the stationary/equilibrium distributions for each of the three datasets, and what accounts for these differences?

There are certainly significant differences in the stationary distributions for each dataset (pictured below).

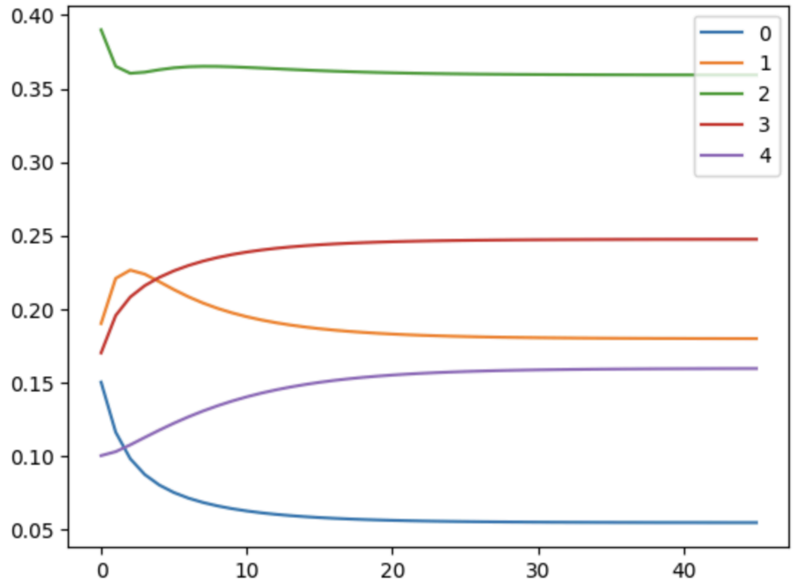
Amnesty International:

```
array([0.16621093, 0.28496106, 0.28423949, 0.16948412, 0.0951044 ])
```



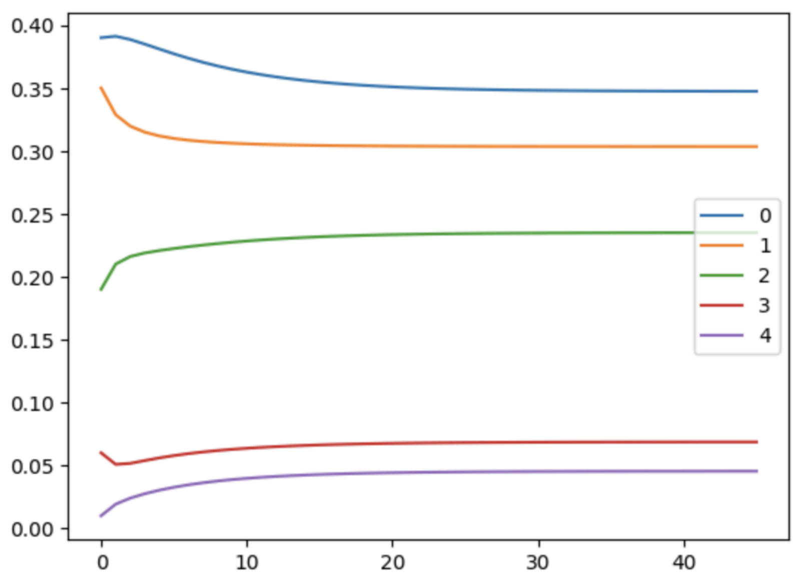
Human Rights Watch:

```
array([0.05433624, 0.17961645, 0.35922214, 0.24740358, 0.1594216 ])
```



State Department:

```
array([0.3473215 , 0.30347185, 0.23505696, 0.06869516, 0.04545452])
```



It was expected that the stationary distributions for Amnesty and HRW would be more similar than when compared to the distribution for the State Department, because the AI and HRW are similar in structure and function (both human rights organizations), and would therefore likely have similar interests and reporting tactics. In contrast, the State Department is not a human rights organization and instead has interests that promote the United States.

Upon closer inspection, however, it appears that all equilibrium distributions are quite different in their own respects. There are no two matrices that are significantly more similar to each than the other. In fact, even though they are quite different, the stationary distributions for Amnesty International and the State Department are more similar to each other than Amnesty is similar to HRW's distribution. One possible explanation for this is the considerable difference in the amount of PTS scoring HRW has than the other two organizations. While the latter have been publishing reports to be scored since 1976, HRW only began participating since 2013, which means there is only eight to nine years worth of data. Thus, the stationary distribution for HRW's scores may not have been reached yet, whereas the 45 years of data on the other entities means that near equilibrium has likely been reached.

If 45 years worth of data was available for HRW, maybe the stationary distributions for HRW and AI would have been more similar. However, it also does not seem correct to conclude a pattern of similarity between AI and the State Department simply because the distributions are "more" similar. Their respective stationary distributions are still significantly different enough to raise the question of discrepancies in reporting and interests between both, since all reports are coded uniformly.

Can this model be used as a sound predictor of future score distributions?

There is value in using these Markov chains to predict future score distributions (and the stationary distribution), however, there are distinctions as to which chains are more useful and reliable in terms of both pragmatism and accuracy. For example, it would likely be better to use the individual profile of models for each dataset rather than the averaged version, because PTS scores themselves are not averaged and instead coded individually. Therefore, predicting scoring for subsequent years should be made using each individual model for better predictability.

Looking into the accuracy of the scores (not the accuracy of the predictions, but whether the scores reflect the actual state of the country), it would be wiser to use the Markov chains pertaining to Amnesty International or HRW because its interests lie in honest reporting more than the State Department's. On prediction accuracy, the State Department and AI are likely the optimal options because there is much more data upon which the chains were created. Because HRW has significantly less data that has gone into constructing the Markov chain, prediction accuracy may be lower as that system has probably not entered into some sort of a stationary distribution yet.

One issue, however, lies in the fundamental property of a Markov chain in that history is completely discarded in prediction. If it is safe to operate under the assumption that the history of this system is not integral to predicting the score distributions, then a more confident conclusion to this question would be in the affirmative. Of course, a Markov chain cannot be used to judge whether history actually matters for the PTS system. Therefore, some form of hypothesis testing that would conclude whether history is independent of score distributions would likely guide the verdict on the predictability of the model.

Are there any interventions that could change the course of the stationary distributions permanently?

Human rights groups, many governments, and other groups have the larger goal of reducing repression (and in effect, the PTS scores of these highly repressive states), but as noted above, it is, in fact, highly repressive states that tend to be stuck in the same state after every interval. Enacting interventions to permanently alter the stationary distributions so that more countries face low political terror would mean acting to alter the transition probabilities themselves. However, there are several challenges with this, because even if the solution is obvious (i.e. democratization, education), the implementation always faces the same road block every time: sovereignty.

Democracies with free and fair elections are often associated with more respect for human rights and less political terror, yet how can anyone convince a dictator or an oligarchy to relent their power with sovereignty as the established norm? It seems the only way to spread democracy is through some form of uprising or revolt. The risk for democratization could then be large-scale violence and terror in the process of achieving this goal, because it is often such revolutions that topple dictatorships to establish democracies often result in violence perpetrated mostly by the state as a form of retaliation and preservation of the status quo. So, a country may see a score distribution of {4, 4, 5, 5, 5, 4, 4, 3, 2, 2} , where the revolution was staged after the second interval. Can the rights violations during the revolution be justified? In other words, do the ends justify the means? Perhaps not, because still, there are no guarantees that democracy will flourish after successfully overthrowing a repressive regime, as history has proven time and time again. Power struggles continue, and repression, although in alternate forms, can remain (Egypt is an example of this). Still, it is important to note that nonviolent uprisings by civilians to topple repressive regimes have higher success rates than violent movements, though the success rate itself is typically dismal.

Thus, the interventions themselves to alter these transition probabilities are mostly known, but successful, and equally importantly, peaceful, implementation that remains a mystery. Sovereignty is an extremely crucial principle, yet ironically, remains an obstacle to reducing repression, especially state-perpetrated repression.

Limitations, Implications and Future Areas for Investigation

Although it was previously noted that liberal democracies are associated with lesser repression while their authoritarian counterparts are associated with the high levels of repression, this is not necessarily always the case, and is important to note some interesting exceptions to this trend. Total dictatorships, or authoritarian regimes that have a history of extreme repression, often find themselves receiving lower PTS scores than partial democracies. For example, Qatar consistently receives a low score, and while it is true that political terror is scarce, the state is classified non-democratic. Qatar is an overall wealthy and small country, with a mostly homogenous population. These factors are all correlated with being less repressive, because the government does not have to use political terror as a tactic to suppress its population- people tend to agree with and abide by the rule of law. Another example would be China, which has consistently received a lower PTS score on average than democracies like India, despite being significantly less democratic than India. Scholars attribute this to China's history of extreme repression that has, in effect, scared people into being obedient, and this obedience continues well into the modern age. Therefore, the government has no reason to employ political terror, and the environment is relatively peaceful. In contrast, India has historically been classified as an illiberal democracy, in which although there are generally free and fair elections, other characteristics such as freedom of speech tend to be more regulated by the government, and repression is enacted to suppress populations that are difficult to control diplomatically. This classification reflects a higher PTS score for India than China, thus posing a limitation on the interpretation of PTS score distribution given regime type.

Moreover, as noted above, the dilemma of history still stands. It is difficult to conclude with certainty whether the history of a state's score distribution is significant in predicting its future scores, which the Markov chain does not take into account by nature. To remedy this, perhaps the current model can be combined or used with a model that does account for history, such as Polya's Urn model. This may be able to explain why countries that have the highest and lowest scores tend to remain in those states as time goes on, because given the "urn," which would represent a country's score distribution, a ball (modeling the PTS score) would be drawn, and returned to the urn with another ball of the same score. If this model is applied to this scenario, a pattern could emerge in which each country that fits the case of a history of extreme scores could have an urn that is composed mostly of "balls" of one score, either 1 or 5.

Markov chains are certainly useful for analyzing the convergence of the distributions to equilibrium and demonstrating the probabilities of moving between scores given the initial scores for a country, but cannot necessarily explain why such probabilities exist, or why equilibrium manifests in a certain way. The most useful application of the model would likely be in its ability to predict the future distribution of a system.

Further investigation should definitely be conducted into being able to better explain the discrepancies between the Markov chains for each entity- are the differences significant enough to conclude that there is strong bias present? If so, how are these biases harmful in the model's ability to accurately predict the political terror within a state? If HRW and Amnesty International are similar in purpose and motive, then why are their stationary distributions noticeably different from one another? Is the lack of data for HRW a valid enough explanation for this?

It also must not be forgotten that it is not practical for these systems to reach a stationary distribution in reality, though they may approach something similar. This is because one of the assumptions of Perron-Frobenius Theorem, fixed transition probabilities, do not hold in reality, and are subject to change at any given moment across repeated intervals. To further complicate the current model, perhaps another Markov chain that modeled the probability of transition states changing given the score could be constructed to account for these potential fluctuations, and then somehow combined with the original Markov chains.

Finally, there is also potential to extend the Markov chains in their current form. Perhaps, improvements upon the prediction accuracy of the Markov Model can be accomplished by creating a model similar to how random forests are constructed, but for Markov chains. Instead of creating transition probabilities using all the data as one sample, a large number K of random samples sized N > 30 based on year should be used to construct K Markov chains, and then these Markov chains can be aggregated to create a model similar to a random forest, which combines several decision trees to improve prediction accuracy.

Introducing categorization may also make the chains more accurate. Would creating Markov chains categorized by regime type better help explain the macro-trends, and would the stationary distributions be different from each other? Maybe this would be misleading, because countries that proclaim themselves as democratic in principle may be autocratic in practice. How can we account for this? Considering Markov chains in combination with other models, such as threshold models could also be useful for determining why certain regimes are more repressive than others. For example, if the leader of a country sees that two out of three of its neighboring countries are experiencing civil conflict, would it be more likely to employ repression as a tactic to ensure stability within its own borders? Would this circumstance make that country more likely to remain among high PTS score states even after the conflicts dissipate?

To paint a more comprehensive picture that offers deeper insight into why the distributions produced through the models exist, it is important to consider the many-model approach to compensate, in a way, for the limitations of Markov chains.