STATISTICAL METHODS FOR IDENTIFYING PERIODS OF STATE-BASED CONFLICT

Thomas Palma [1] Edward Austin [2]

[1]Lancaster University

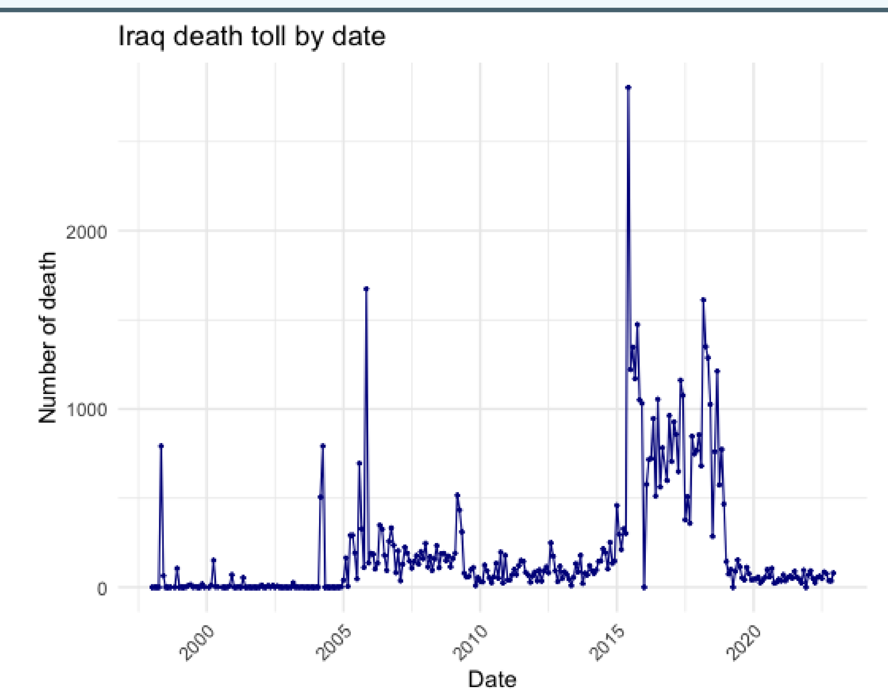
1. Motivation

This project is motivated by the challenge of identifying periods of state-based conflict in a geographical region. The motivation for applying count data time series models to death counts in state-based conflicts is to understand and predict how conflicts evolve over time. This approach helps:

- Characterize Conflict Dynamics
  - Reveals patterns in how violence fluctuates, highlighting periods of escalation and de-escalation.
- Forecast Future Trends
  - Provides predictions on future conflict trends, aiding in proactive intervention and preparedness.
- Support Informed Decision-Making
  - Offers data-driven insights for policymakers to allocate resources and adjust strategies effectively.
- Enhance Conflict Prevention
  - Identifies patterns that inform targeted prevention and response strategies.

2. State-Based Conflict

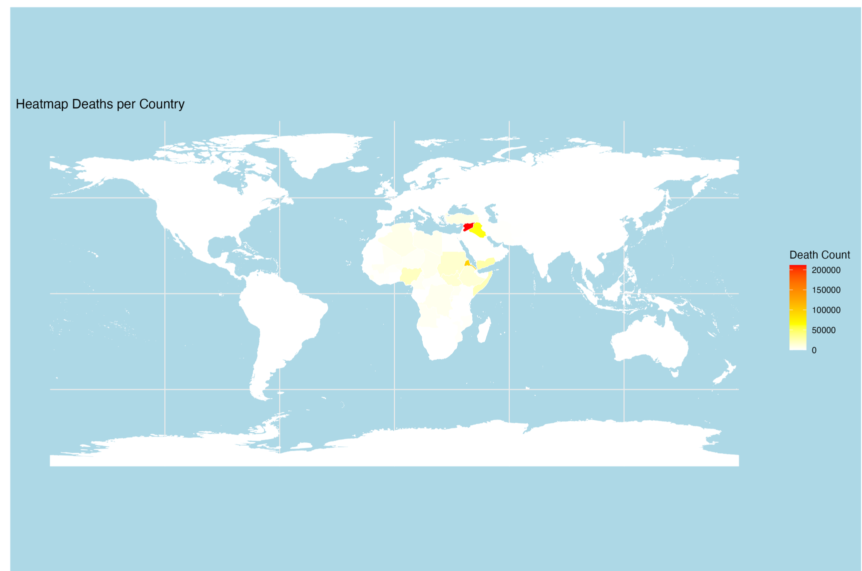
Involves violence where at least one party is a state, including conflicts between states or between a state and rebels/civilians. A year with 25 deaths from such violence is classified as a conflict period. This measure is imprecise for assessing conflict duration and does not confirm if a conflict is ongoing. Timely prediction is crucial for effective intervention.



3. Comments on Data

The dataset on daily death counts in Iraq presents several important characteristics:

- Frequency of Deaths:
  - Most days show zero deaths, suggesting low conflict or reporting gaps. Significant death spikes on specific days may indicate major events or intensified conflict.
- Temporal Trends:
  - Death tolls vary greatly over time, reflecting changes in conflict intensity or strategies. Analyzing these trends can reveal seasonal or cyclical patterns in the conflict.
- Outliers:
  - High death counts, like a peak of 2803 deaths in one day, may signal major incidents or data issues. Investigating these outliers helps understand their impact on overall conflict analysis.



4. Poisson AR

Poisson Distribution:
Models the number of events occurring in a fixed interval with a constant rate. Its PMF is:

$$P(Y=y)=\frac{\lambda^y e^{-\lambda}}{y!}$$

where λ is the average rate and y is the observed count.
Autoregressive Model (AR(1)):
Captures temporal dependence in a time series:

$$Y_t=\phi Y_{t-1}+\epsilon_t$$

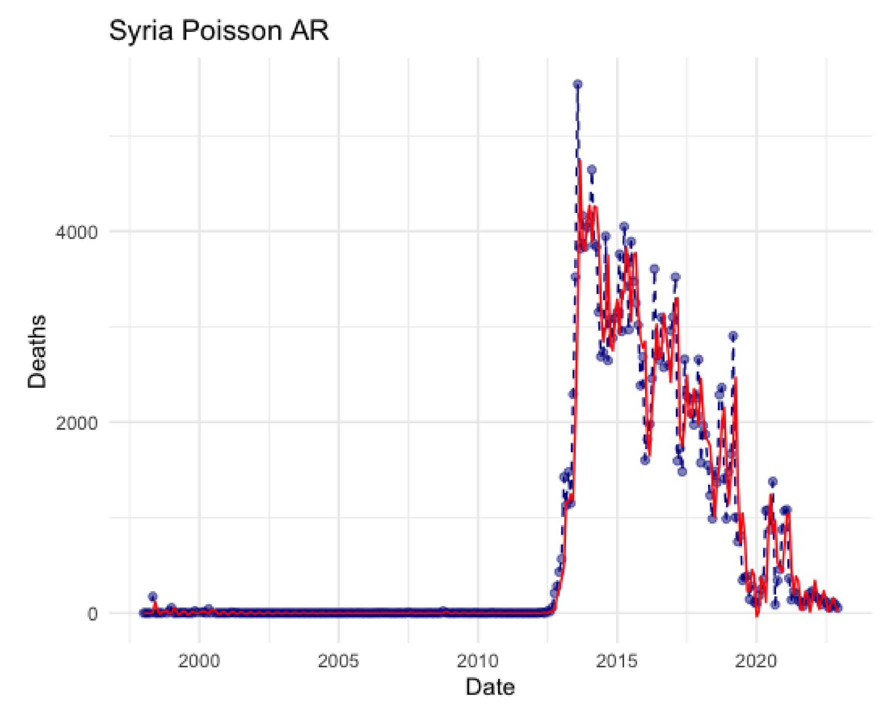
where ϕ is the autoregressive parameter and ϵ_t is the error term.
Poisson AR Model:
Combines Poisson distribution with AR(1) for count data with temporal dependencies. The rate λ_t is:

$$\log(\lambda_t)=\beta_0+\beta_1\text{lag}_t+\beta_2\text{date}_t$$

with:

$$Y_t=\phi Y_{t-1}+\epsilon_t$$

5. Poisson AR Fit



8. Comparison

Poisson AR Model	ZIPAR Model
Handling Excess Zeros Assumes count data follows a Poisson distribution, which struggles with excess zeros.	**Handling Excess Zeros** Handles excess zeros by distinguishing between structural zeros (inherent characteristics) and sampling zeros (high rate of zeros in Poisson process).
Better Fit for Overdispersed Data Assumes mean and variance are equal, leading to poor fit if variance exceeds mean.	**Better Fit for Overdispersed Data** Accounts for overdispersion by separating zero-generating and count-generating processes.
Improved Model Accuracy May inaccurately estimate rate parameter (λ) with many zeros.	**Improved Model Accuracy** Accurately captures data-generating process, leading to better parameter estimates and predictions.
Enhanced Forecasting and Inference Challenging to forecast with many zeros due to inadequate zero-count modeling.	**Enhanced Forecasting and Inference** Models zero-inflation and count processes separately, providing better forecasts and reliable inferences.

6. Zero-Inflated Poisson AR

The ZIP distribution accounts for excess zeros using a mixture of two components:

$$P(Y=y)=\begin{cases}\pi+(1-\pi)e^{-\lambda}, & \text{if } y=0\\ (1-\pi)\frac{\lambda^y e^{-\lambda}}{y!}, & \text{if } y>0\end{cases}$$

where π is the probability of excess zeros and λ is the Poisson rate.

Autoregressive Model (AR(1)):
Captures temporal dependency:

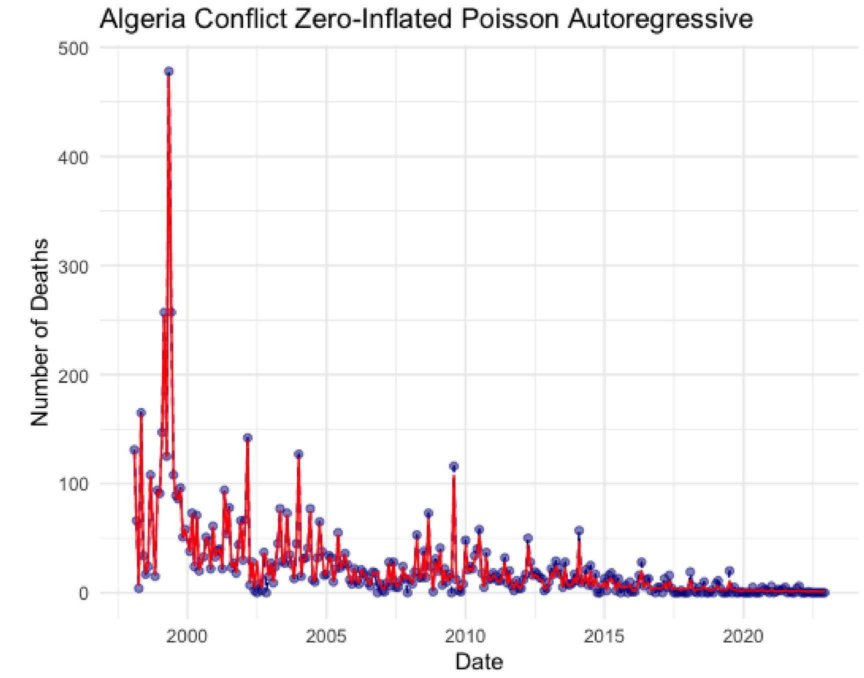
$$Y_t=\phi Y_{t-1}+\epsilon_t$$

where ϕ is the autoregressive parameter and ϵ_t is the error term.

ZIP-AR Model:

- **Count Process:**
$$\log(\lambda_t)=\beta_0+\beta_1\text{lag}_t+\beta_2\text{date\_num}_t$$
- **Zero-Inflation:**
$$\text{logit}(\pi_t)=\gamma_0+\gamma_1\text{date\_num}_t$$
- **Random Effects:**
$$\text{random effect}\sim\mathcal{N}(0,\sigma^2)$$

7. Zero-Inflated Poisson AR Fit



9. Future Work

For future work, the application of the negative binomial distribution to model state-based conflict data is a promising direction. This approach is particularly suitable for over-dispersed data, where the variance exceeds the mean, which is often the case with conflict-related death counts.

- **Overdispersion:** The negative binomial distribution is advantageous over the Poisson distribution, which assumes equal mean and variance. The added flexibility of the negative binomial model makes it more appropriate for datasets with high variability.
- **Model Comparison:** Future research could involve fitting the negative binomial model to historical conflict data and comparing its performance with other models, such as Poisson and zero-inflated Poisson models. This comparison could provide insights into which model best captures the nature of conflict data.
- **Policy Implications:** By providing more accurate predictions of conflict periods, this approach can help policymakers intervene more effectively and allocate resources where they are most needed.

References

- Liboschik, Tobias, Kerschke, Pascal, and Fried, Roland. *tscount: An R package for analysis of count time series following generalized linear models*. Journal of Statistical Software, 2017. Vol. 82, No. 5, pp. 1-51.
- Hyndman, Rob J., and Athanasopoulos, George. *Forecasting: Principles and Practice*. OTexts, 2018. 2nd edition.

t.palmachaundler@lancaster.ac.uk