Statistical Methods for Identifying Periods of State-Based Conflict

Thomas Palma[1]
Supervised by Ed Austin[2]

[1]Lancaster University
[2]STOR-i Centre for Doctoral Training

30 of August, 2024

Introduction

- **State-based conflict** involves violence where at least one party is a government, including conflicts between states or between the government and rebels or civilians.
- Conflicts often involve civilians, making it crucial for decision-makers to identify periods of conflict early to intervene appropriately.
- The project aims to apply statistical methods to identify and analyze periods of state-based conflict.

Motivation

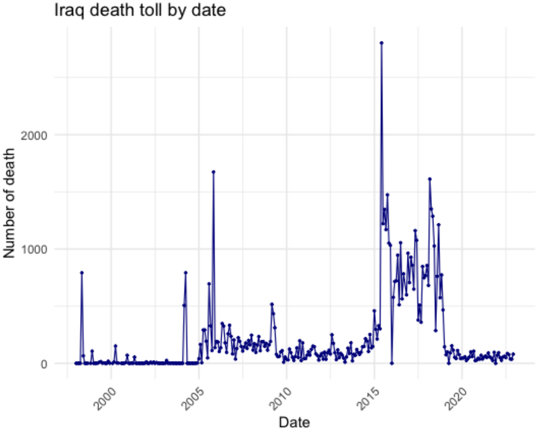
- Understanding and predicting how conflicts evolve over time is essential for proactive intervention.
- The project uses count data time series models to characterize conflict dynamics and forecast future trends.
- Insights from this analysis can support informed decision-making and enhance conflict prevention strategies.

State-Based Conflict Definition

- Defined as violence where at least one party is the government of a state, including conflicts between states or between a government and rebels/civilians.
- A year with 25 deaths from such violence is classified as a conflict period.
- This definition may hinder precise statistical modeling of conflict duration and intensity.
- Prediction of conflict periods is crucial for timely intervention.

Example - Iraq Conflict

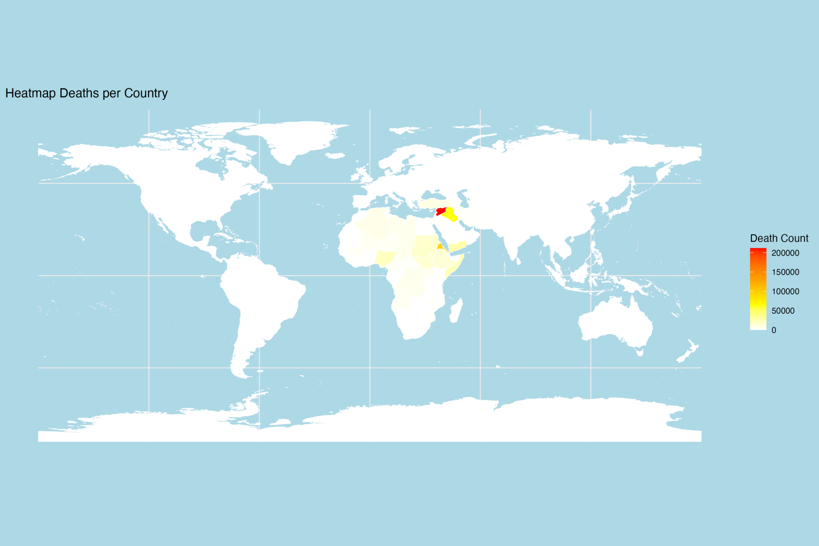
- Iraq is a prominent example of a state-based conflict, with frequent clashes involving the government and various insurgent groups.
- The conflict in Iraq provides rich data for statistical analysis, demonstrating both the intensity and duration of such conflicts.



Comments on Data

- **Frequency of Deaths:** Most days show zero deaths, with spikes indicating major events.
- **Temporal Trends:** Death tolls vary greatly over time, reflecting changes in conflict intensity or strategies.
- **Outliers:** High death counts, such as a peak of 2803 deaths in one day, may signal major incidents or data issues.

Heatmap


Heatmap Deaths per Country
Death Count
200000
150000
100000
50000
0

Model Comparison Process

To compare time series models, follow these steps:

1. **Simulate Data**: Generate data from a chosen time series model.
2. **Fit Models**: Fit different models to the simulated data.
3. **Record Metric**: For each model, record the performance metric, such as the Sum of Squared Errors (SSE):

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

4. **Repeat**: Repeat this process over K iterations (e.g., 500), recording the SSE for each iteration.
5. **Compute Mean**: Calculate the Mean Squared Error (MSE) for each model:

$$\mathrm{MSE}^{(m)} = \frac{1}{K} \sum_{k=1}^{K} \sum_{t=1}^{T} (y_{t,k} - \hat{y}_{t,k}^{(m)})^2$$

Poisson AR Model

Poisson Distribution:

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

where λ is the average rate and y is the observed count.

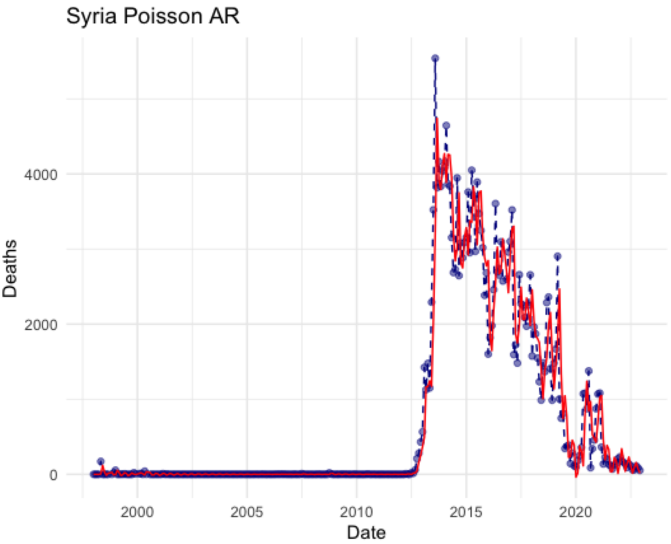
Autoregressive Model (AR(1)):

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

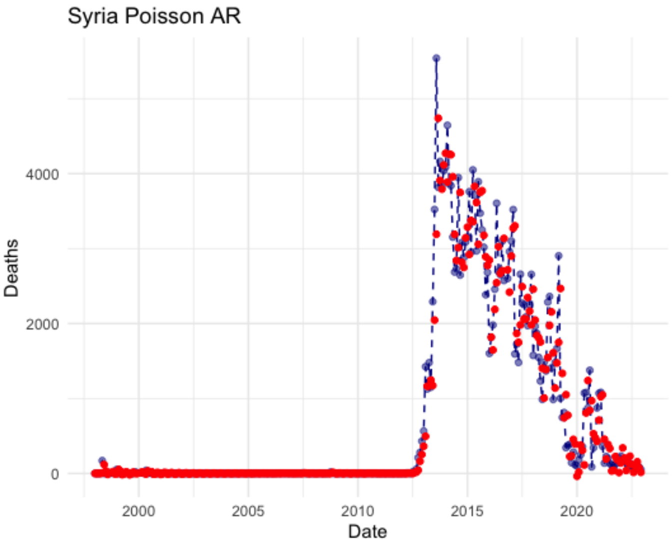
where ϕ is the autoregressive parameter and ϵ_t is the error term.

Poisson AR Model: Combines Poisson distribution with AR(1) for count data with temporal dependencies.

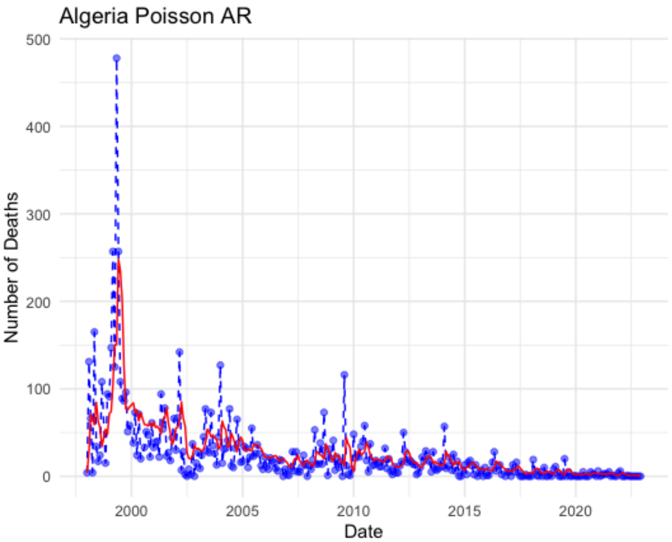
Poisson AR Fit



Poisson AR Fit



Bad fits


Algeria Poisson AR
500
400
300
200
100
0
Number of Deaths
2000
2005
2010
2015
2020
Date

Zero-Inflated Poisson AR Model

ZIP Distribution:

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

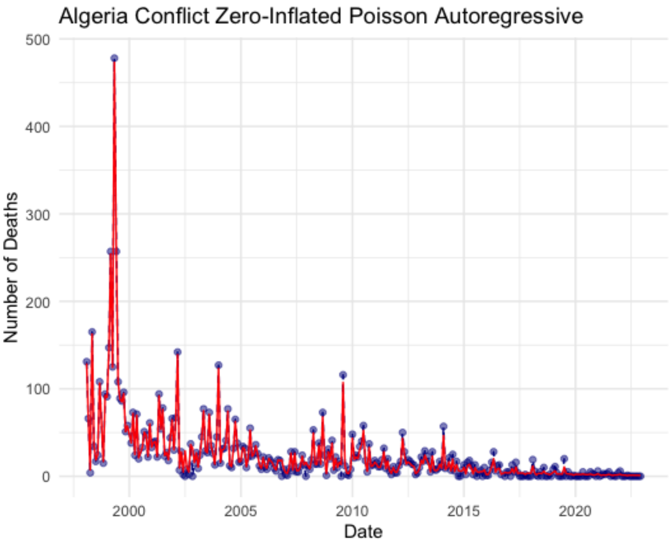
where π is the probability of excess zeros and λ is the Poisson rate.

ZIP-AR Model:

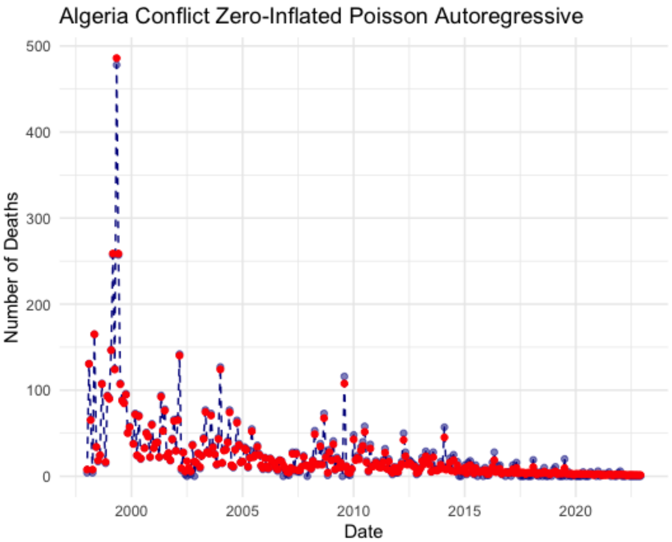
$$\log(\lambda_t) = \beta_0 + \beta_1 \text{lag}_t + \beta_2 \text{date}_t$$

$$\text{logit}(\pi_t) = \gamma_0 + \gamma_1 \text{date}_t$$

Zero-Inflated Poisson AR Fit



Zero-Inflated Poisson AR Fit



Comparison: Handling Excess Zeros

Poisson AR Model

Handling Excess Zeros
Assumes count data follows a Poisson distribution, which struggles with excess zeros.

ZIPAR Model

Handling Excess Zeros
Handles excess zeros by distinguishing between structural zeros (inherent characteristics) and sampling zeros (high rate of zeros in Poisson process).

Comparison: Fit for Overdispersed Data

Poisson AR Model

Worst Fit for Overdispersed Data
Assumes mean and variance are equal, leading to poor fit if variance exceeds mean.

ZIPAR Model

Better Fit for Overdispersed Data
Accounts for overdispersion by separating zero-generating and count-generating processes.

Comparison: Model Accuracy

Poisson AR Model

Model Accuracy
May inaccurately estimate rate parameter (λ) with many zeros.

ZIPAR Model

Model Accuracy
Accurately captures data-generating process, leading to better parameter estimates and predictions.

Comparison: Forecasting and Inference

Poisson AR Model

Forecasting and Inference
Challenging to forecast with many zeros due to inadequate zero-count modeling.

ZIPAR Model

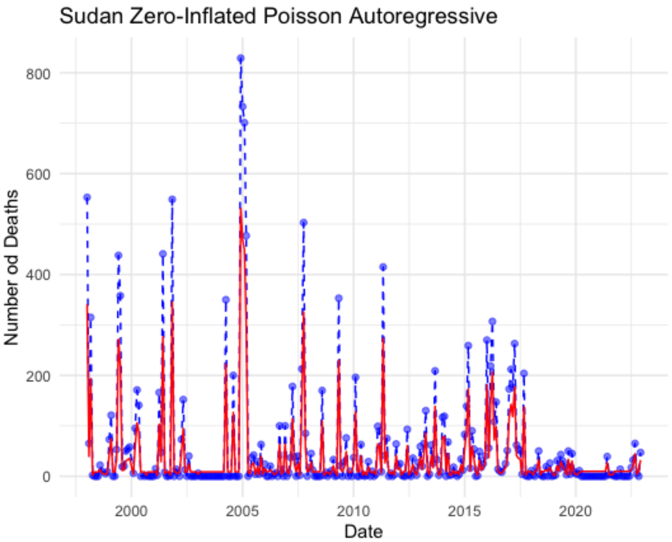
Forecasting and Inference
Models zero-inflation and count processes separately, providing better forecasts and reliable inferences.

Future Work: Overdispersion

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.

Bad fits



Sudan Zero-Inflated Poisson Autoregressive
800
600
400
200
0
Number od Deaths
2000
2005
2010
2015
2020
Date

Future Work: Model Comparison

- **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.

References

- Petersson et al. (2019). "State-Based Conflict Analysis."
- Hyndman, R. J. (2018). "Forecasting Principles and Practice."