

Did you know that a hormone Oxytocin that plays important role in reproduction, caring and social behavior was found to cause people trust more strangers?

Think about potential monetizing opportunity...

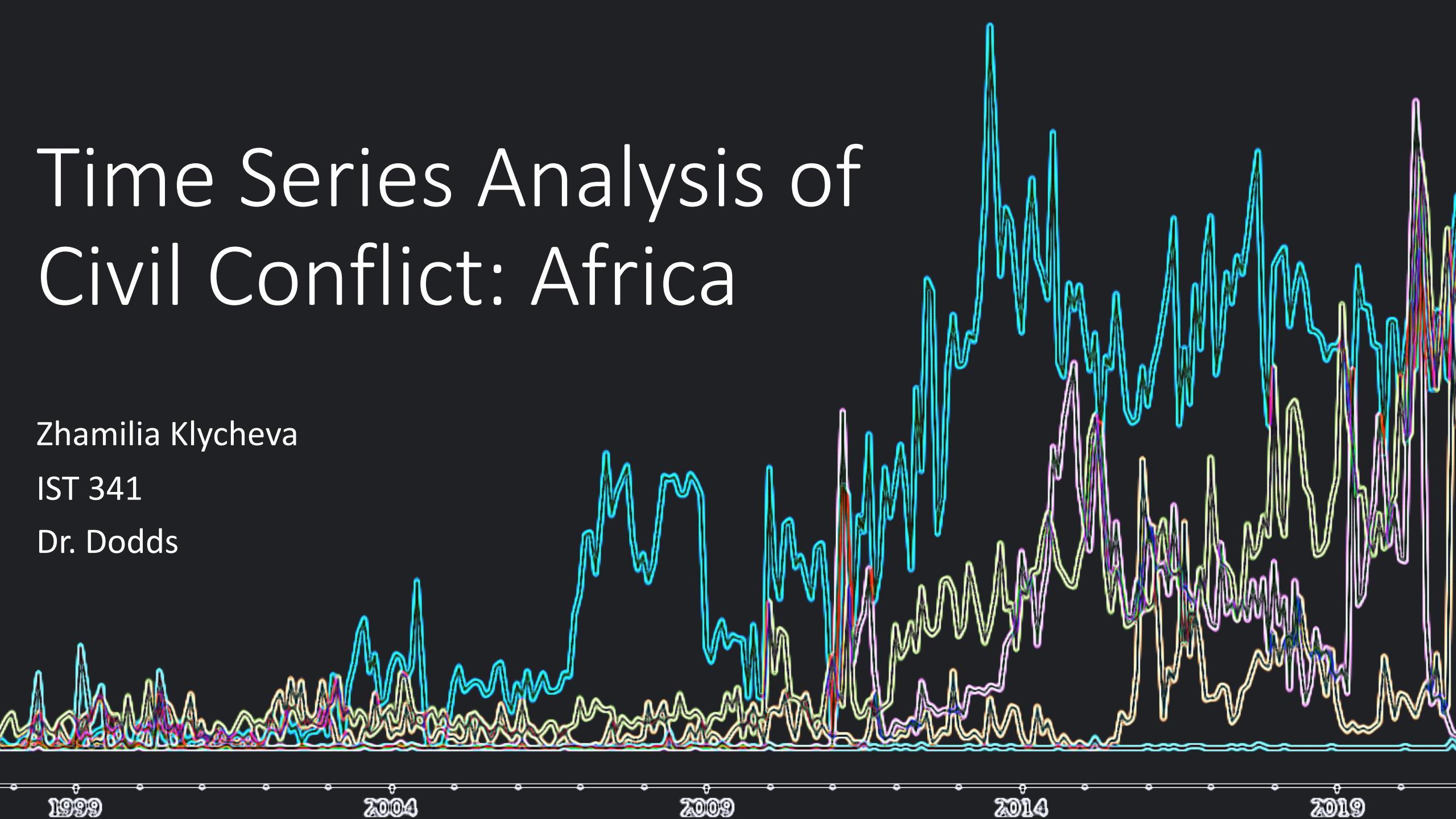


Time Series Analysis of Civil Conflict: Africa

Zhamilia Klycheva

IST 341

Dr. Dodds



Motivation/Big Picture problem

- Study of Civil Conflict
- Some ongoing conflicts in Africa
 - Ethiopia
 - Libya
 - Nigeria
 - Somalia
- How can we predict conflicts based on prior events?

Conflict Event Data

- Battles
- Explosions/remote violence
- Violent Protests
- Riots
- Violence Against civilians
- Time period: 1997-Present
- 292 observations
- Source: ACLED

Libraries/resources

- Numpy
- Pandas
- Matplotlib - plots
- Datetime – working with date, collapsing data
- Sklearn
- **statsmodels.tsa**
 - Stattools
 - Arima
 - tsaplots

Progress thus far

- Data cleaning



- Exploratory Data Analysis



- ARIMA Modeling (Box Jenkins Approach) - Ethiopia



Data Format: Time Series

| somaliaEvent_date | somalia_year | somalia_event | somalia_fatalities |
|-------------------|--------------|---------------|--------------------|
| 1997-02-28 | 1997 | 1 | 1 |
| 1997-03-31 | 1997 | 1 | 0 |
| 1997-04-30 | 1997 | 1 | 2 |
| 1997-05-31 | 1997 | 0 | 0 |
| 1997-06-30 | 1997 | 10 | 39 |

Object of interest: Somalia

Unit of analysis: month

Data Cleaning

| event_date | data_id | iso | event_id_no_cnty | year | time_precision | inter1 | inter2 | interaction | admin3 | l |
|------------|---------|-----|------------------|------|----------------|--------|--------|-------------|--------|-----|
| 1997-02-28 | 7797591 | 706 | 28750 | 1997 | | 3 | 2 | 7 | 27 | 0.0 |
| 1997-03-31 | 6875883 | 706 | 1 | 1997 | | 1 | 3 | 3 | 33 | 0.0 |
| 1997-04-30 | 7534426 | 706 | 2 | 1997 | | 1 | 3 | 7 | 37 | 0.0 |
| 1997-05-31 | 0 | 0 | 0 | 1997 | | 0 | 0 | 0 | 0 | 0.0 |

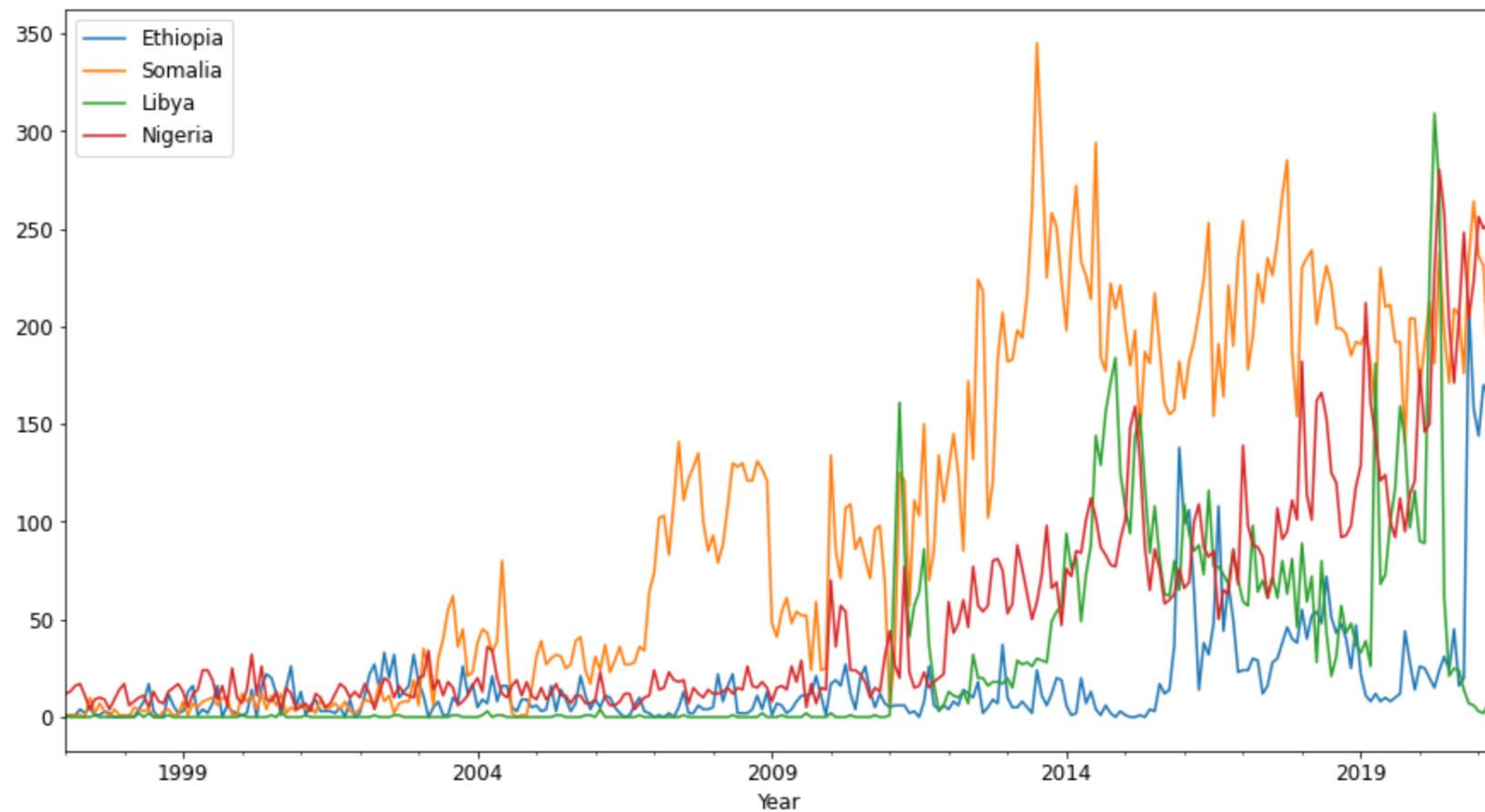
Each row: individual event occurrence

Remove all “non-violent” events

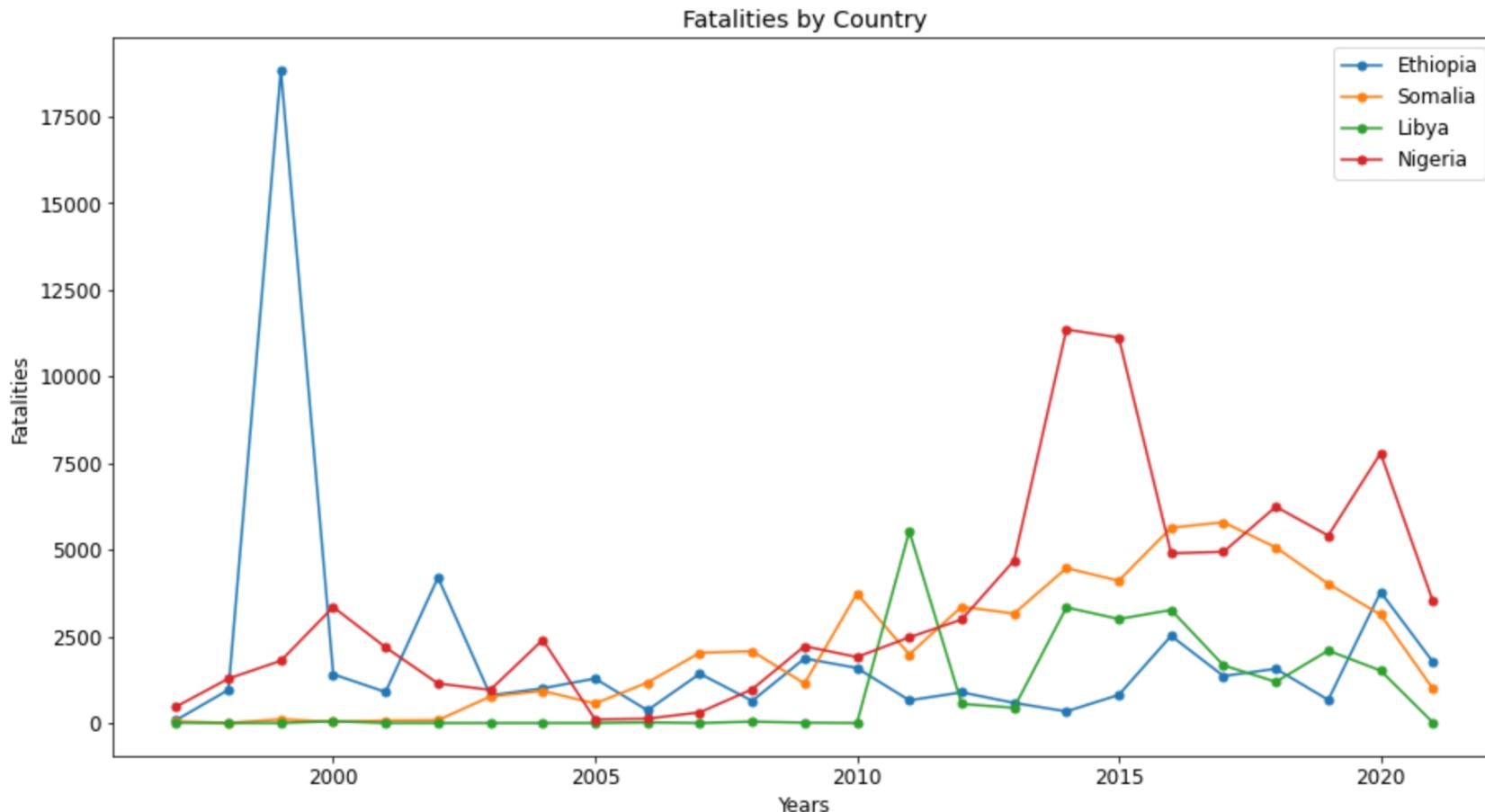
| somaliaEvent_date | somalia_year | somalia_event | somalia_fatalities |
|-------------------|--------------|---------------|--------------------|
| 1997-02-28 | 1997 | 1 | 1 |
| 1997-03-31 | 1997 | 1 | 0 |
| 1997-04-30 | 1997 | 1 | 2 |
| 1997-05-31 | 1997 | 0 | 0 |
| 1997-06-30 | 1997 | 10 | 39 |

Group by month & count number
of events per month

EDA: Conflict Time Series by Country



EDA: Fatalities Time Series by Country



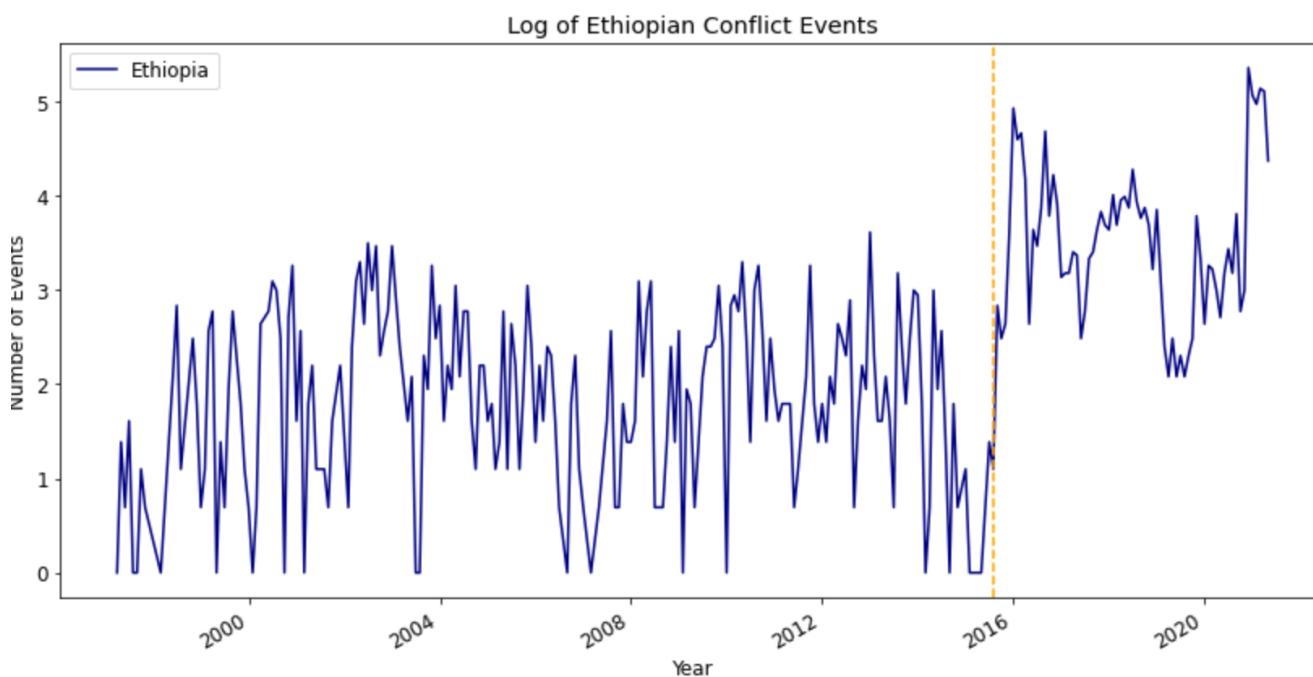
Box-Jenkins Approach

1. Model Identification
 - a. Autocorrelation, Partial Autocorrelation
 - a. Parameter Estimation
 - a. MLE, OLS
 - b. Statistical Model checking
 - a. Residual plot, Ljung-Box test, ACF/PACF of residuals

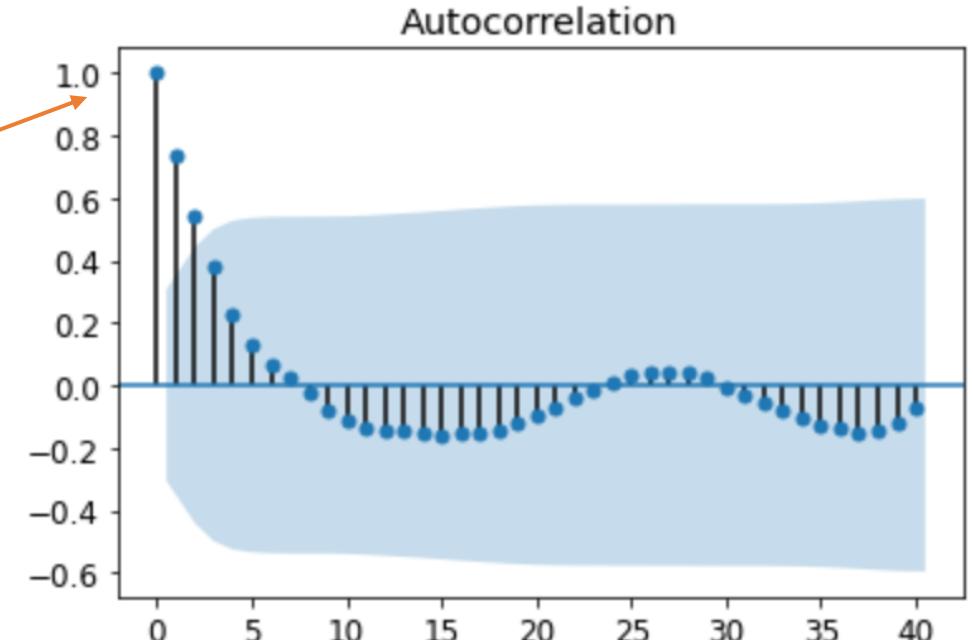
Ethiopia



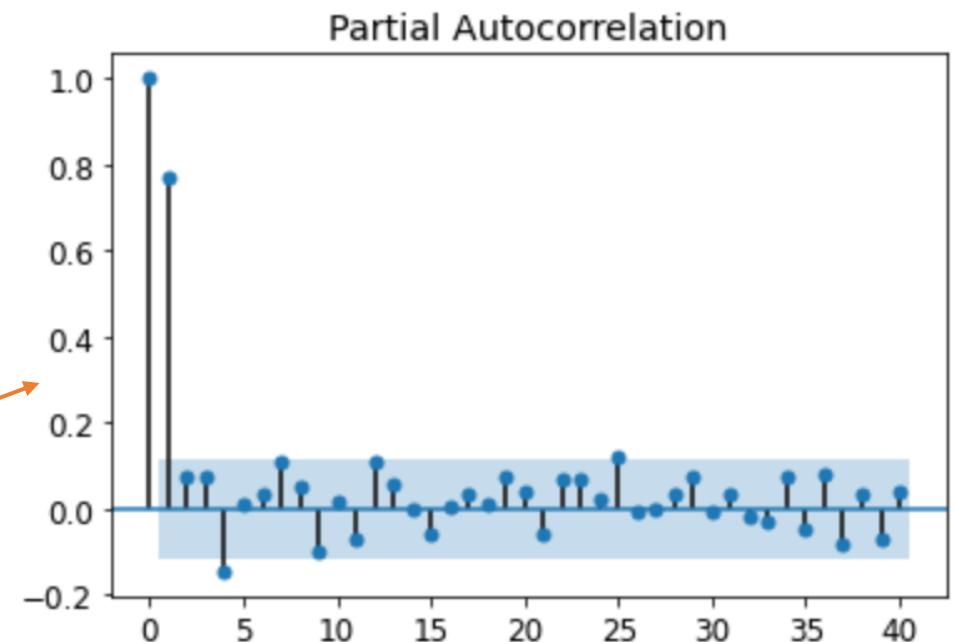
1. Model Identification



Helps determine the number of lags
 q for Moving Average (MA)



Helps determine the number of lags
 p for Autoregressive Process (AR)



2-3. Parameter Estimation/ ARIMA Model Evaluation

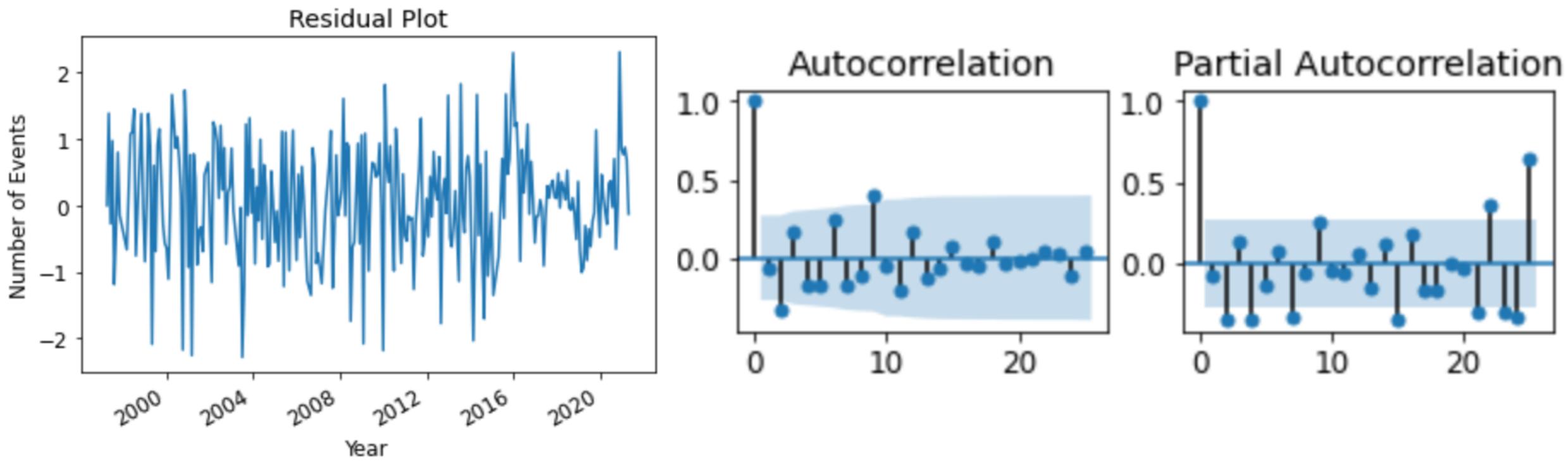
| ARIMA (p,d,q) | Log Likelihood | AIC | BIC | HQIC |
|---------------|----------------|---------|---------|---------|
| ARIMA (1,1,1) | -339.850 | 685.701 | 696.485 | 690.032 |
| ARIMA (2,1,1) | -339.772 | 687.543 | 701.922 | 693.318 |
| ARIMA (2,1,2) | -335.539 | 681.077 | 699.051 | 688.296 |
| ARIMA (1,1,2) | -339.631 | 687.262 | 701.640 | 693.036 |
| ARIMA (1,1,0) | -360.410 | 724.819 | 732.009 | 727.707 |
| ARIMA (2,1,0) | -350.458 | 706.917 | 717.701 | 711.247 |

Sample ARIMA (1,1,1) Output, look at the significance of parameters

| SARIMAX Results | | | | | | |
|-------------------------|------------------|-------------------|----------|-------|--------|--------|
| Dep. Variable: | log_ethiopia | No. Observations: | 270 | | | |
| Model: | ARIMA(1, 1, 1) | Log Likelihood | -339.850 | | | |
| Date: | Fri, 14 May 2021 | AIC | 685.701 | | | |
| Time: | 09:44:00 | BIC | 696.485 | | | |
| Sample: | 0 - 270 | HQIC | 690.032 | | | |
| Covariance Type: | opg | | | | | |
| | coef | std err | z | P> z | [0.025 | 0.975] |
| ar.L1 | 0.3782 | 0.073 | 5.208 | 0.000 | 0.236 | 0.521 |
| ma.L1 | -0.8810 | 0.044 | -19.831 | 0.000 | -0.968 | -0.794 |
| sigma2 | 0.7303 | 0.065 | 11.235 | 0.000 | 0.603 | 0.858 |
| Ljung-Box (Q): | 30.26 | Jarque-Bera (JB): | 2.15 | | | |
| Prob(Q): | 0.87 | Prob(JB): | 0.34 | | | |
| Heteroskedasticity (H): | 0.73 | Skew: | -0.22 | | | |
| Prob(H) (two-sided): | 0.13 | Kurtosis: | 3.06 | | | |

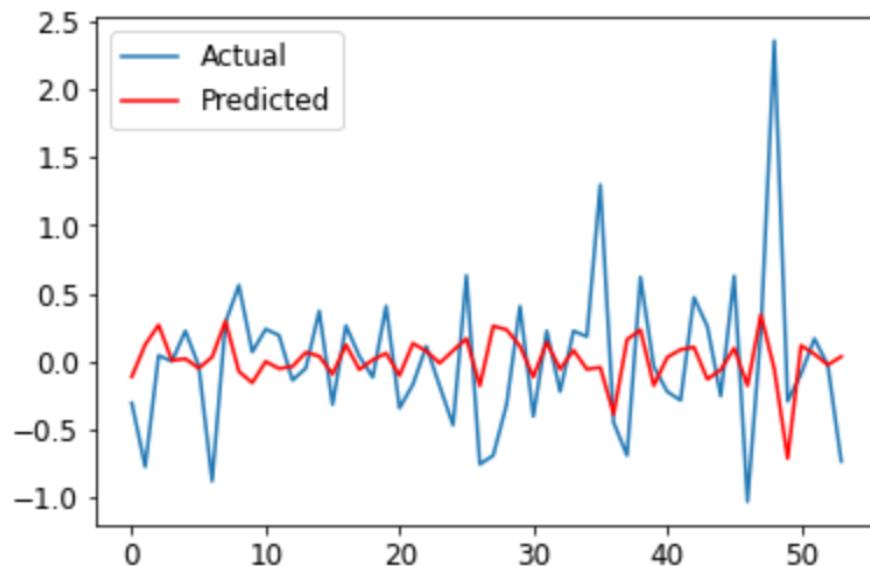
3. Statistical Model checking

ARIMA (1,1,1)



ARIMA Rolling Forecast

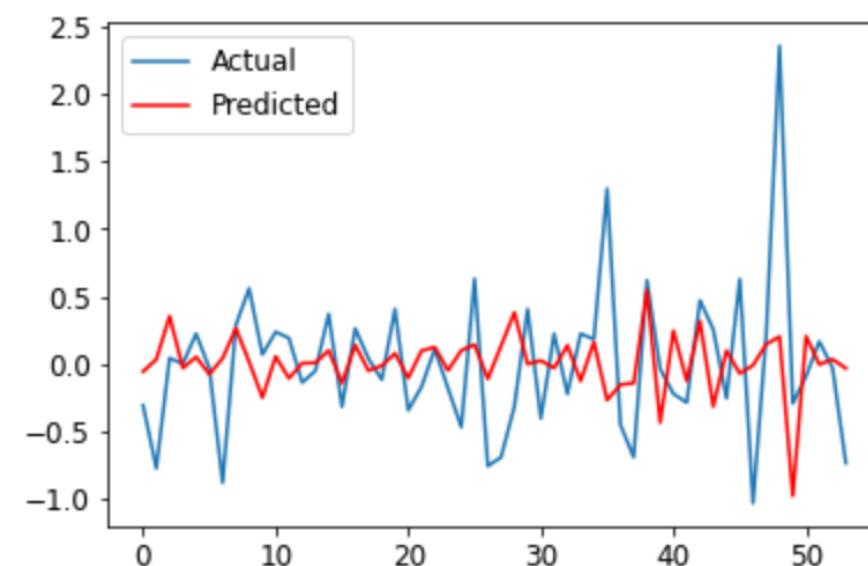
ARIMA (1,1,1)



Train-Test: 80/20

Test RMSE: 0.547

ARIMA (2,1,2)

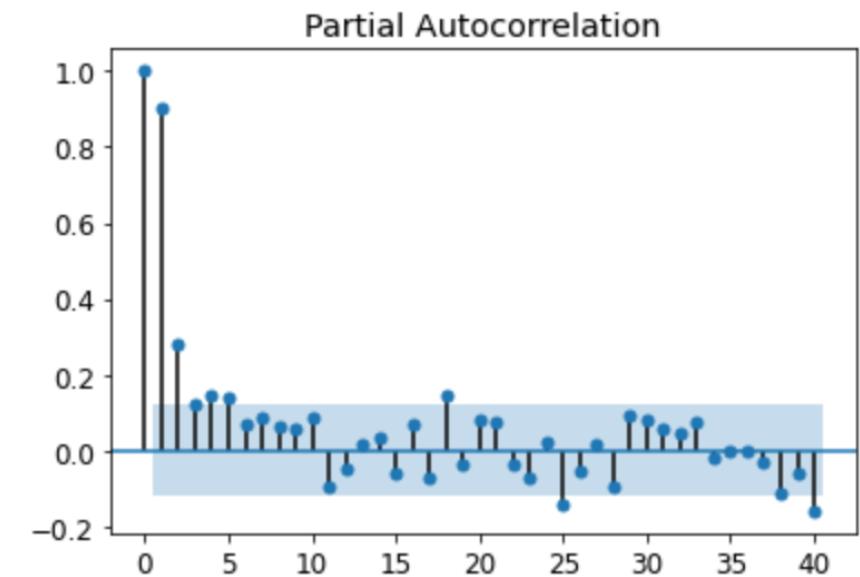
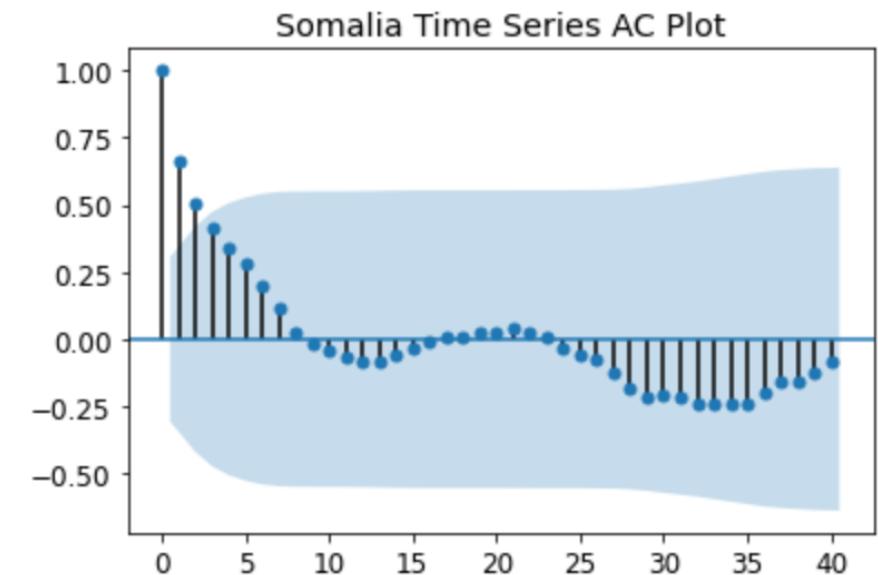
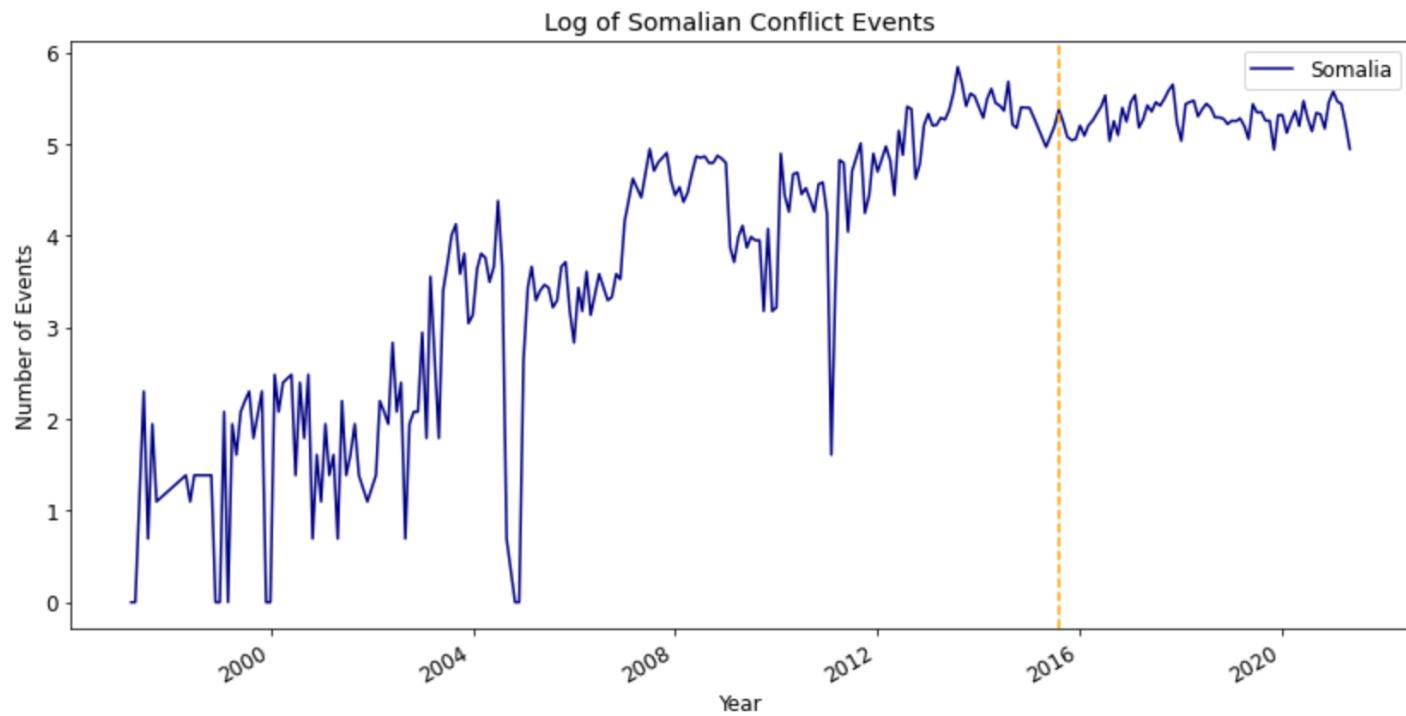


Train-Test: 80/20

Test RMSE: 0.553

Somalia

1. Model Identification

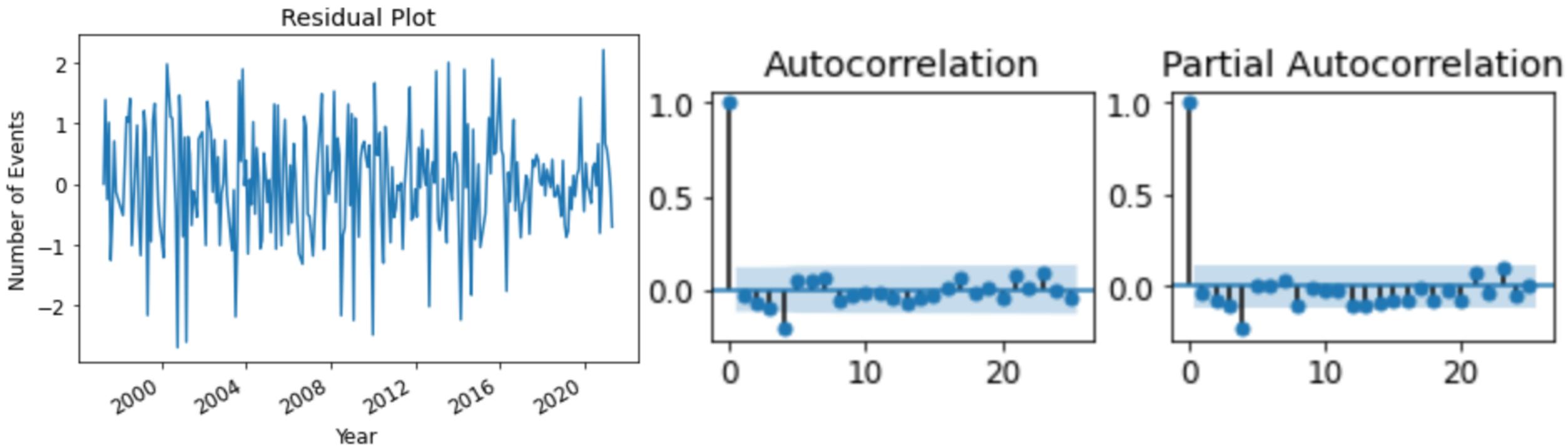


2-3. Parameter Estimation/ ARIMA Model Evaluation

| ARIMA(p,d,q) | Log Likelihood | AIC | BIC | HIC |
|--------------|----------------|---------|---------|---------|
| ARIMA(3,1,1) | -333.927 | 681.853 | 707.016 | 691.959 |
| ARIMA(2,1,2) | -335.539 | 681.077 | 699.051 | 688.296 |
| ARIMA(2,1,1) | -339.772 | 687.543 | 701.922 | 693.318 |
| ARIMA(2,1,0) | -350.458 | 706.917 | 717.701 | 711.247 |
| ARIMA(3,1,0) | -348.265 | 704.529 | 718.908 | 710.304 |
| ARIMA(1,1,0) | -360.410 | 724.819 | 732.009 | 727.707 |

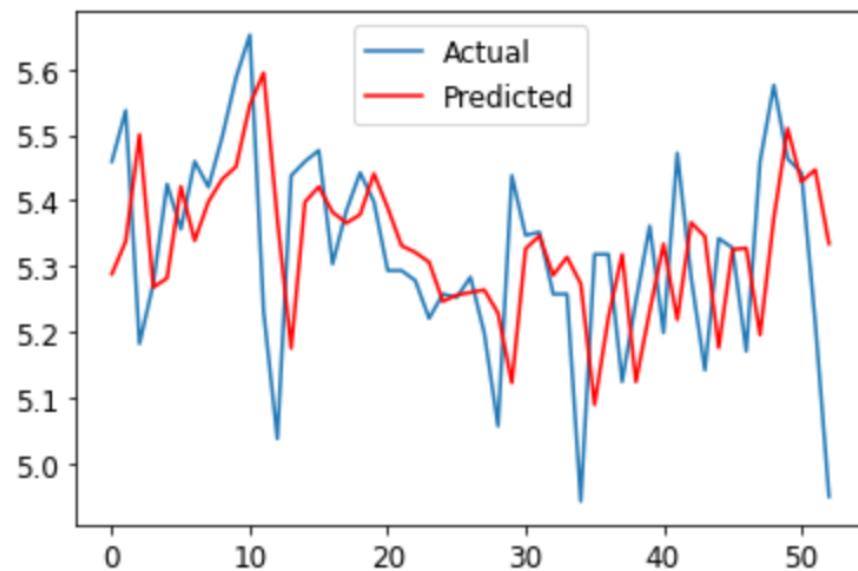
3. Statistical Model checking

ARIMA (3,1,0)



ARIMA Rolling Forecast

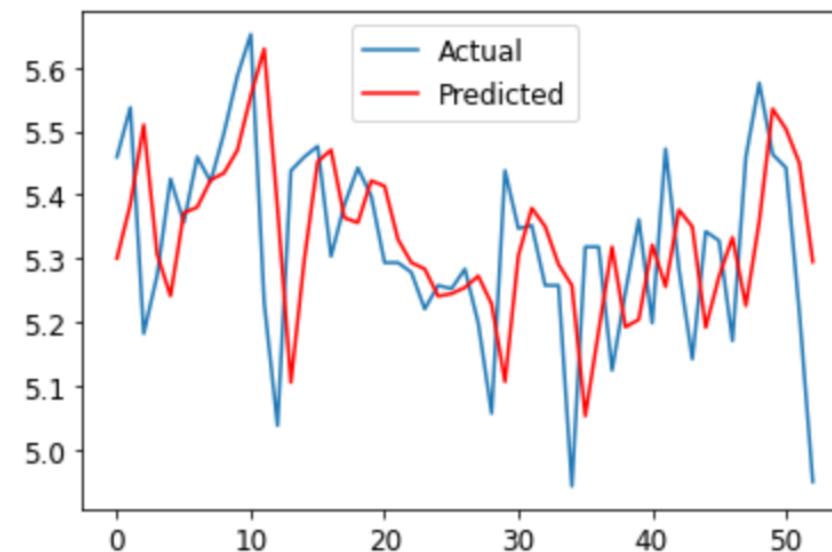
ARIMA (3,1,0)



Train-Test: 80/20

Test RMSE: 0.167

ARIMA (1,1,0)



Train-Test: 80/20

Test RMSE: 0.172

Conclusion

- Ethiopia TS data found ARIMA(1,1,1) as the best fit
- Somalia TS data found ARIMA(3,1,0) as the best fit
- ARIMA (3,1,0) for Somalia had a higher forecasting power than ARIMA(1,1,1) for Ethiopia (the logic is taking the best fitting models for each country and evaluating based on out-of-sample forecasting)
- Overall, very interesting to work with the event data as each country is unique and requires different model
- ARIMA has shown to be effective in fitting the data