

ts_xterisation_rca

October 17, 2019

Time Series Characterisation

Table of Contents

- 1 Time Series
- 2 Importing Libraries for time series forecasting
- 3 Importing data
- 4 Data Preprocessing and Visualization
 - 4.1 Stationarity
 - 4.1.1 ACF and PACF plots
 - 4.1.2 Plotting Rolling Statistics
 - 4.1.3 Augmented Dickey-Fuller Test
 - 4.2 Making Time Series Stationary
 - 4.2.1 Transformations
 - 4.2.1.1 Log Scale Transformation
 - 4.2.1.2 Other possible transformations:
 - 4.2.2 Techniques to remove Trend - Smoothing
 - 4.2.2.1 Moving Average
 - 4.2.2.2 Exponentially weighted moving average:
 - 4.2.3 Further Techniques to remove Seasonality and Trend
 - 4.2.3.1 Differencing
 - 4.2.3.2 Decomposition

DATA SOURCES

This project makes use of a time series of export data from **The Observatory of Economic Complexity(OEC)** based on the SITC (Standard International Trade Classification)** from 1962 - 2000, with data from **The Center for International Data** from Robert Feenstra. The more recent data 2001 - 2017, is sourced from **UN COMTRADE**.

Full **OEC** trade datasets are also available from a data dump on **SITC Product 4 Digit**

1 Importing Libraries for time series forecasting

```
[98]: #Run this for quick setup  
#!/usr/bin/env python3  
#!pip install -r requirements.txt
```

```
[99]: import numpy as np  
import pandas as pd  
pd.set_option('display.expand_frame_repr', False)  
pd.set_option('display.max_columns', 500)  
pd.set_option('display.width', 1000)  
  
import statsmodels.robust  
import statsmodels.tsa.stattools as tsa  
  
from statsmodels import api as sm  
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf  
from statsmodels.tsa.stattools import adfuller  
from statsmodels.tsa.seasonal import seasonal_decompose  
  
from statsmodels.graphics.tsaplots import plot_acf  
from statsmodels.graphics.tsaplots import plot_pacf  
from pylab import rcParams  
  
from math import sqrt  
  
import matplotlib  
import matplotlib.pyplot as plt  
matplotlib.colors  
matplotlib.rcParams['axes.labelsize'] = 14  
matplotlib.rcParams['xtick.labelsize'] = 12  
matplotlib.rcParams['ytick.labelsize'] = 12  
matplotlib.rcParams['text.color'] = 'k'  
import seaborn as sns  
  
from random import random  
  
import warnings; import os  
warnings.filterwarnings("ignore")  
plt.style.available  
plt.style.use('bmh')  
import os
```

2 Importing data

- Dataset: Trade data:
 - location_id: / - product_id: SITC 4 digit code system
 - year: from 1963 to 2017 - export_value: / - import_value: - / - export_rca:

- is_new: / - cog: / - distance: / - normalized_distance: / - normalized_cog:
- normalized_pci: / - export_rpop: / - sitc_eci: / - sitc_coi: / - pci: / - location_code: / - location_name_short_en:
- sitc_product_code: / - sitc_product_name_short_en: / - Unit: arbitrary

```
[100]: product = 'engine_parts' # Create a dictionary to store product code and sitc_code
experiment = 'rca_tests'

def load_data(csv_name, trade_form, target_value, code_system, sitc_code, ts_type):
    '''
        Ingests SITC multi-time series trade data in csv format, filters for one product code, removes zero values for target value, screens target values and renames target columns & creates dataframes for the target columns

    Parameters:
    -----
    csv:
    trade_form:
    target_value:
    code_system:
    sitc_code:
    ts_type:
    '''

    data = pd.read_csv(csv_name) #, header=None
    # TO DO: Create for loop to read sitc code from dictionary
    dframe = data.loc[data[code_system]==sitc_code]
    dframe.rename(columns={'location_name_short_en': 'exporter', 'export_rca': 'rca'}, inplace=True)
    dframe = dframe[[ts_type, trade_form, target_value]]
    dframe = dframe[dframe[trade_form] != 'Undeclared Countries']
    dframe[target_value].replace('', 0, inplace=True)
    dframe = dframe[dframe[target_value] != 0]
    dframe = dframe[dframe[target_value] >= 0.09]
    dframe = dframe[dframe[target_value] <= 20]
    return dframe
```

```
[101]: xdf = load_data('sitc4digit_year.csv', 'exporter', 'rca', 'sitc_product_code', 7132, 'year')
xdf.columns
```

```
[101]: Index(['year', 'exporter', 'rca'], dtype='object')
```

```
[102]: exporters = xdf['exporter'].nunique()
time = xdf['year'].nunique()
unknown = xdf['exporter'].loc[xdf['exporter']=='Undeclared Countries'].count()
```

```
ctotal = xdf.exporter.count()
zerorca = xdf['rca'][xdf['rca']==0].count()
nulls = xdf['rca'].isnull().sum().sum()
```

```
[103]: print(f'Shape of the dataframe: {xdf.shape}')
print(f'The data covers {time} years from {xdf.year.min()} to {xdf.year.max()} for {exporters} countries')
print(f'The rca values range from {xdf.rca.min()} to {xdf.rca.max()}'')
print(f'Countries not declared are {unknown} out of {ctotal}')
print(f'Zero export values: {zerorca}')
print(f'Null export values: {nulls}')
```

```
Shape of the dataframe: (1939, 3)
The data covers 56 years from 1962 to 2017 for 188 countries
The rca values range from 0.09004033 to 17.575996
Countries not declared are 0 out of 1939
Zero export values: 0
Null export values: 0
```

3 Data Preprocessing and Visualization

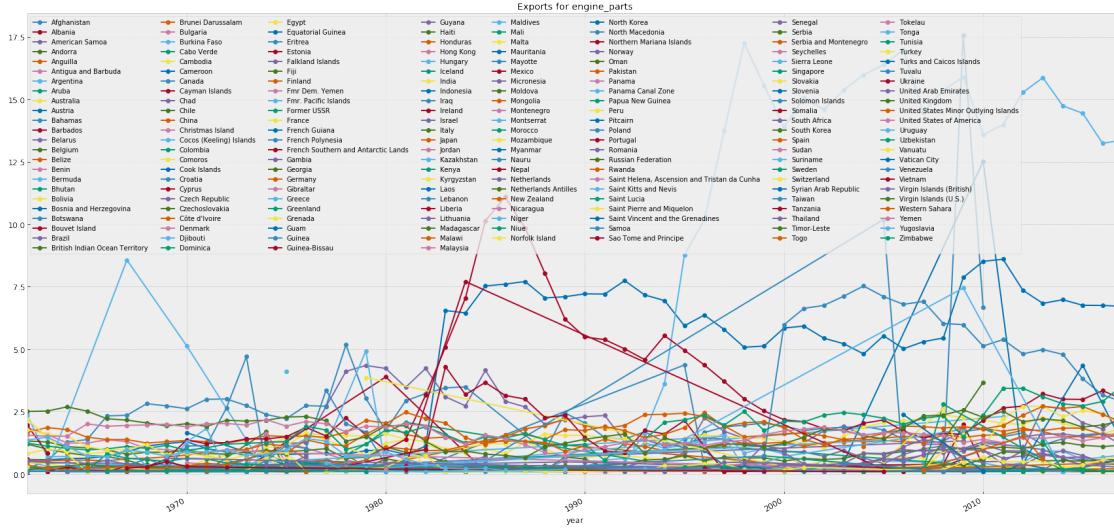
Converting to datetime format:

```
[104]: xdf['year'] = pd.to_datetime(xdf['year'], format='%Y')
xdf.columns
```

```
[104]: Index(['year', 'exporter', 'rca'], dtype='object')
```

```
[105]: #xdf = xdf.drop(columns=["rca"])
#xdf.columns
```

```
[107]: fig, ax = plt.subplots(figsize=(30,15))
for key, grp in xdf.groupby(['exporter']):
    ax = grp.plot(ax=ax, kind='line', x='year', y='rca', label=key, marker='o', title='Exports for {}'.format(product))
    ax.legend(ncol=8, loc='best')
    target = f'images/{product}/{experiment}'
    savefile = os.path.join(os.getcwd(), target, f'Exports for {product}')
plt.savefig(savefile)
```



```
[156]: xdf.describe(include='all')
```

```
[156]:
```

| | year | exporter | rca |
|--------|---------------------|----------|-------------|
| count | 1939 | 1939 | 1939.000000 |
| unique | 56 | 188 | NaN |
| top | 1975-01-01 00:00:00 | Spain | NaN |
| freq | 54 | 56 | NaN |
| first | 1962-01-01 00:00:00 | NaN | NaN |
| last | 2017-01-01 00:00:00 | NaN | NaN |
| mean | NaN | NaN | 1.193243 |
| std | NaN | NaN | 2.062163 |
| min | NaN | NaN | 0.090040 |
| 25% | NaN | NaN | 0.175712 |
| 50% | NaN | NaN | 0.513809 |
| 75% | NaN | NaN | 1.408146 |
| max | NaN | NaN | 17.575996 |

A majority of countries have rca values below the mean value of 0.636 . A few countries have once-off or intermittently high values preceeded by long periods with zero. These countries were excluded from the dataset and treated as outliers by limiting rca values to below 20, which was observed as the value most countries fall under. An example is Norfolk Islands shown below

```
[109]: rdf = xdf.set_index('year')
#Norfolk = pd.DataFrame(rdf['rca'].loc[rdf['exporter']=='Norfolk Island'])
#Norfolk.to_csv('norfolk.csv')
norfolk = pd.read_csv('norfolk.csv')
norfolk.head()
```

```
[109]:      year      rca
0  1990-01-01  114.697680
1  2001-01-01    0.125032
2  2005-01-01    0.526648
3  2008-01-01   2.614124
4  2009-01-01   0.350387
```

```
[110]: df_grps = xdf.set_index('year')
df_grps.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 1939 entries, 2010-01-01 to 1973-01-01
Data columns (total 2 columns):
exporter    1939 non-null object
rca         1939 non-null float64
dtypes: float64(1), object(1)
memory usage: 45.4+ KB
```

```
[162]: df_grps = xdf.groupby([xdf.year.name, xdf.exporter.name]).mean().unstack()
df_grps = df_grps.apply(pd.to_numeric, errors='coerce').fillna(0, u
    ↪downcast='infer')
df_grps.columns = df_grps.columns.droplevel()
```

Some countries have changed names since 1962, and these names had to be updated along with the data.

```
[163]: df_grps = df_grps.rename(columns={'Fmr. Pacific Islands': 'Pacific Islands'})
df_grps['Russia'] = df_grps['Former USSR']+df_grps['Russian Federation']
df_grps['Yemen'] = df_grps['Fmr Dem. Yemen']+df_grps['Yemen']
```

```
[164]: df_grps = df_grps.drop(['Fmr Dem. Yemen','Norfolk Island','Former
    ↪USSR','Russian Federation','Virgin Islands (U.S.)','Virgin Islands
    ↪(British)'],axis=1)
#df_grps.loc[:, df_grps.any()]
#df_grps.T[(df_grps !=0).any()].T
#df.T[(df != 0).any()].T
```

```
[165]: def remove_dup_cols(frame):
    keep_names = set()
    keep_icols = list()
    for icol, name in enumerate(frame.columns):
        if name not in keep_names:
            keep_names.add(name)
            keep_icols.append(icol)
    return frame.iloc[:, keep_icols]
```

```
[166]: #remove_dup_cols(df_grps)
```

Setting index as the datetime column for easier manipulations:

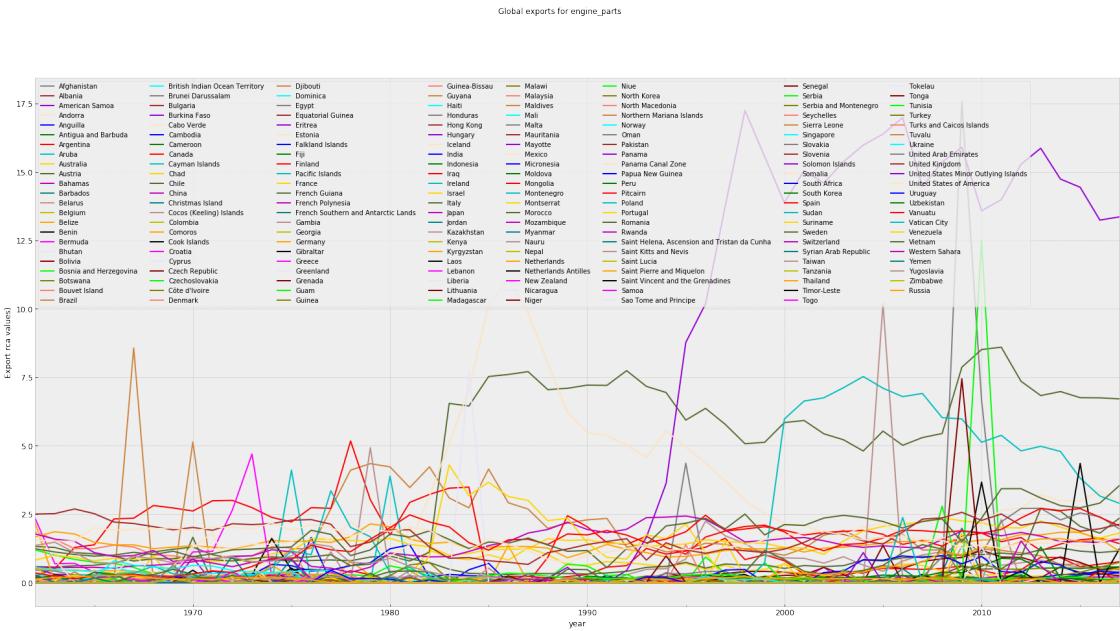
Image storage

```
[167]: location = os.getcwd()
target = f'images/{product}/{experiment}'
path = os.path.join(location, target)

[168]: #df_pivoted = xdf.
        ↪pivot_table(index='year', columns='exporter', values='export_value', aggfunc='sum')
fig1, ax1 = plt.subplots(figsize=(30,15))
#https://matplotlib.org/3.1.0/gallery/color/named_colors.html
colors = □
        ↪["gray", "brown", "darkviolet", "bisque", "b", "g", "r", "c", "gold", "darkolivegreen", "m", "teal", "r",
           "olive", "salmon", "peru", "aqua"]
label='global trends in export rca values for engine parts'
df_grps.plot(ax=ax1,label=label,color=colors)
ax1.set_ylabel('Export rca values')
fig1.suptitle('Global exports for {}'.format(product))

plt.legend(ncol=8,loc='best')

root = os.path.join(path, 'export rca trends')
plt.savefig(str(root))
#.plot(legend=False)#subplots=True, legend=False)
```



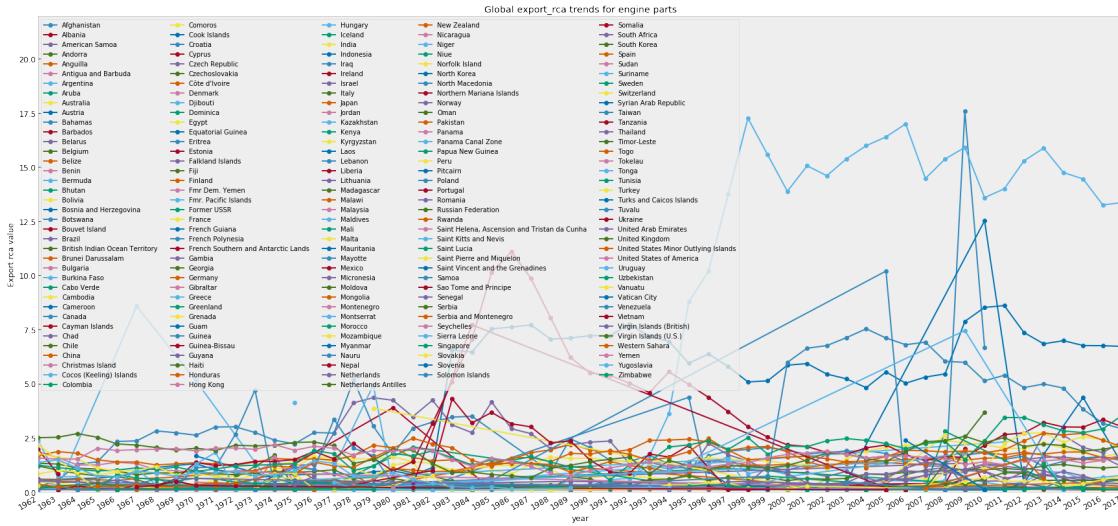
```
[169]: # Adapted from https://stackoverflow.com/questions/30942755/
    ↪plotting-multiple-time-series-after-a-groupby-in-pandas
# Modified from https://www.programcreek.com/python/example/98021/matplotlib.
    ↪dates.YearLocator
from matplotlib.dates import YearLocator, MonthLocator, DateFormatter

def plot_multiple_time_series(dframe, ts_label, grp_label, values_label, □
    ↪product, experiment=experiment, figsize=(30,15), title=None):
    """
    Plots multiple time series on one chart by first grouping time series based
    ↪on series names from column in dataframe

    Parameters:
    -----
    dframe : timeseries Pandas dataframe
    ts_label : string
        The name of the df column that has the datetime timestamp x-axis values.
    grp_label : string
        The column name in dframe for groupby method.
    values_label : string
        The column name in dframe for the y-axis.
    figsize : tuple of two integers
        Figure size of the resulting plot, e.g. (20, 7)
    title : string
        Optional title
    ...
    xtick_locator = YearLocator()
    xtick_dateformatter = DateFormatter('%Y')
    fig, ax = plt.subplots(figsize=figsize)
    location = os.getcwd()
    target = f'images/{product}/{experiment}'
    path = os.path.join(location, target)
    for key, grp in dframe.groupby([grp_label]):
        ax = grp.plot(ax=ax, kind='line', x=ts_label, y=values_label, □
    ↪label=key, marker='o')
        ax.xaxis.set_major_locator(xtick_locator)
        ax.xaxis.set_major_formatter(xtick_dateformatter)
        ax.autoscale_view()
        ax.legend(ncol=5, loc='upper left')
        #_ = plt.xticks(rotation=0, )
        _ = plt.grid()
        _ = plt.xlabel('year')
        _ = plt.ylabel('Export rca value')
        _ = plt.ylim(0, dframe[values_label].max() * 1.25)
        if title is not None:
            _ = plt.title(title)
    savefile = os.path.join(path,label)
```

```
plt.savefig(str(savefile))
_ = plt.show()
```

[170]: `plot_multiple_time_series(xdf, 'year', 'exporter', 'rca',
 ↪'engine_parts', experiment, title='Global export_rca trends for engine parts')`



Reviewing plots of the density of observations will provide further insight into the structure of the data: - Is the distribution perfectly Gaussian (normal distribution)? - Skewness: which direction is the distribution? - If not Gaussian and skew, transformations will be necessary prior to modelling

[171]: `exporters = list(df_grps.columns)`

Target folder to store individual images

```
[172]: def save_images(ts_groups, analysis):
    ...
    analysis: can be from these options
    ↪['decompose', 'acf_pacf', 'adf_test', 'rolling_stats']
    ...
    path = os.getcwd() + f'/images/{product}/{experiment}/{analysis}'
    savefile = os.path.join(path, ts_groups[exporter].name)
    return savefile
```

Plot of histograms, kernel density estimation (KDE) and distribution plots to show skewness in data for each country

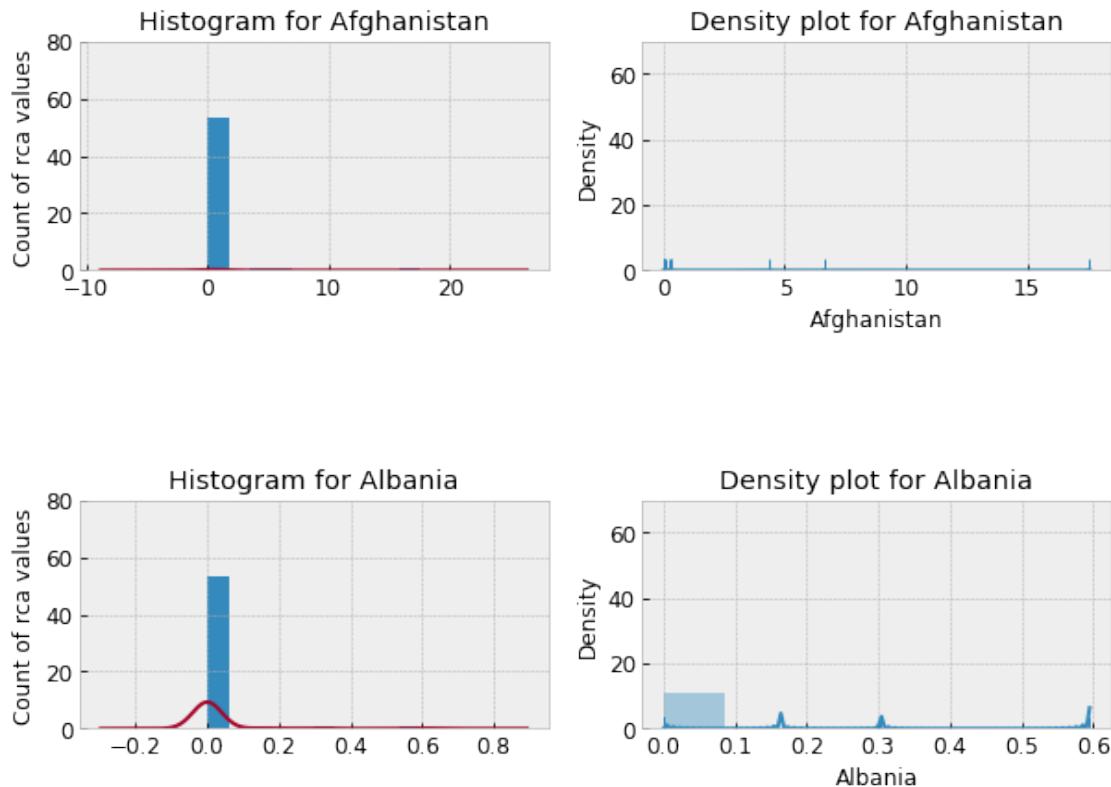
```
[173]: for exporter in exporters:
    root = save_images(df_grps, 'rca_histograms')
    plt.figure(figsize=(10,5));
```

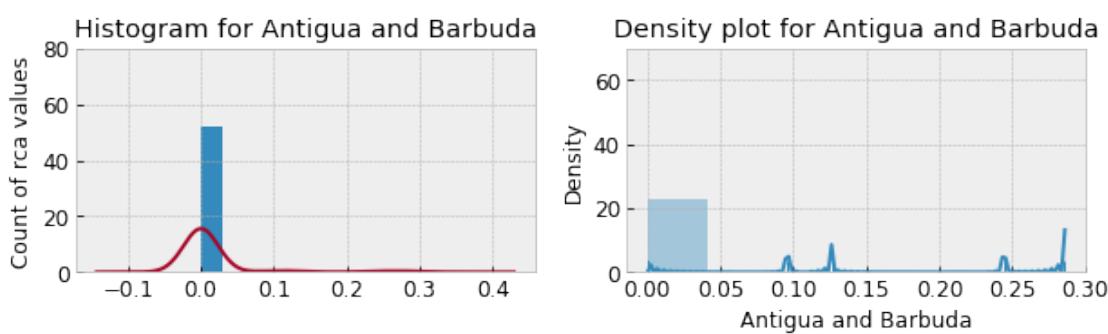
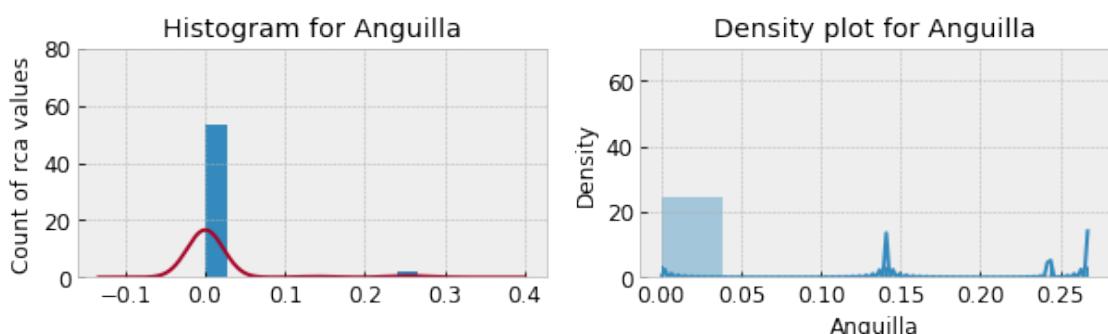
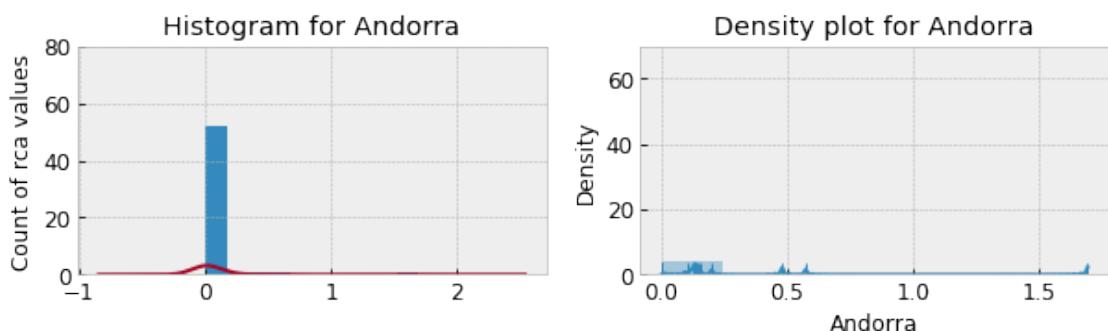
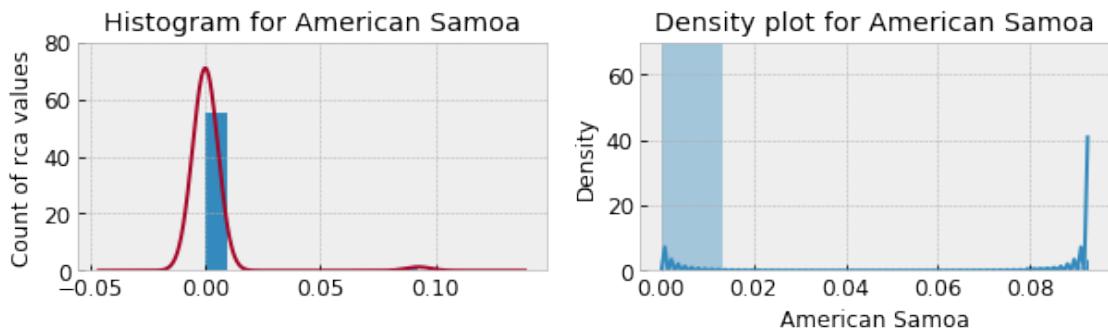
```

plt.subplot(221)
plt.title(f'Histogram for {exporter}')
df_grps[exporter].plot(kind='hist')
df_grps[exporter].plot(kind='kde')
plt.ylabel('Count of rca values')
#plt.xlim(1962,2016)
plt.ylim(0,80)

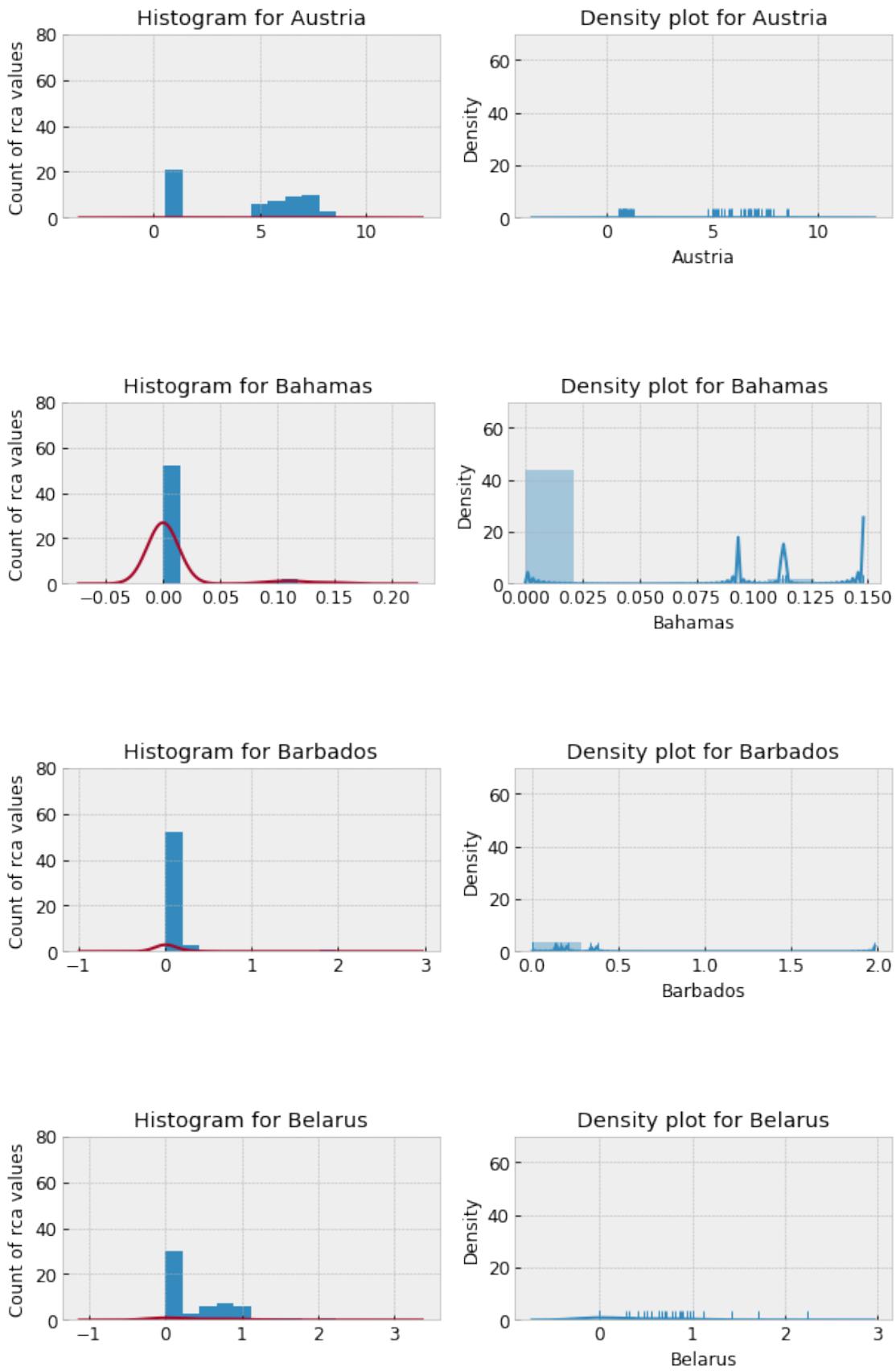
plt.subplot(222)
plt.title(f'Density plot for {exporter}')
sns.distplot(df_grps[exporter],hist=True,rug=True)
plt.ylabel('Density')
plt.ylim(0,70)
plt.savefig(str(root))
plt.show()

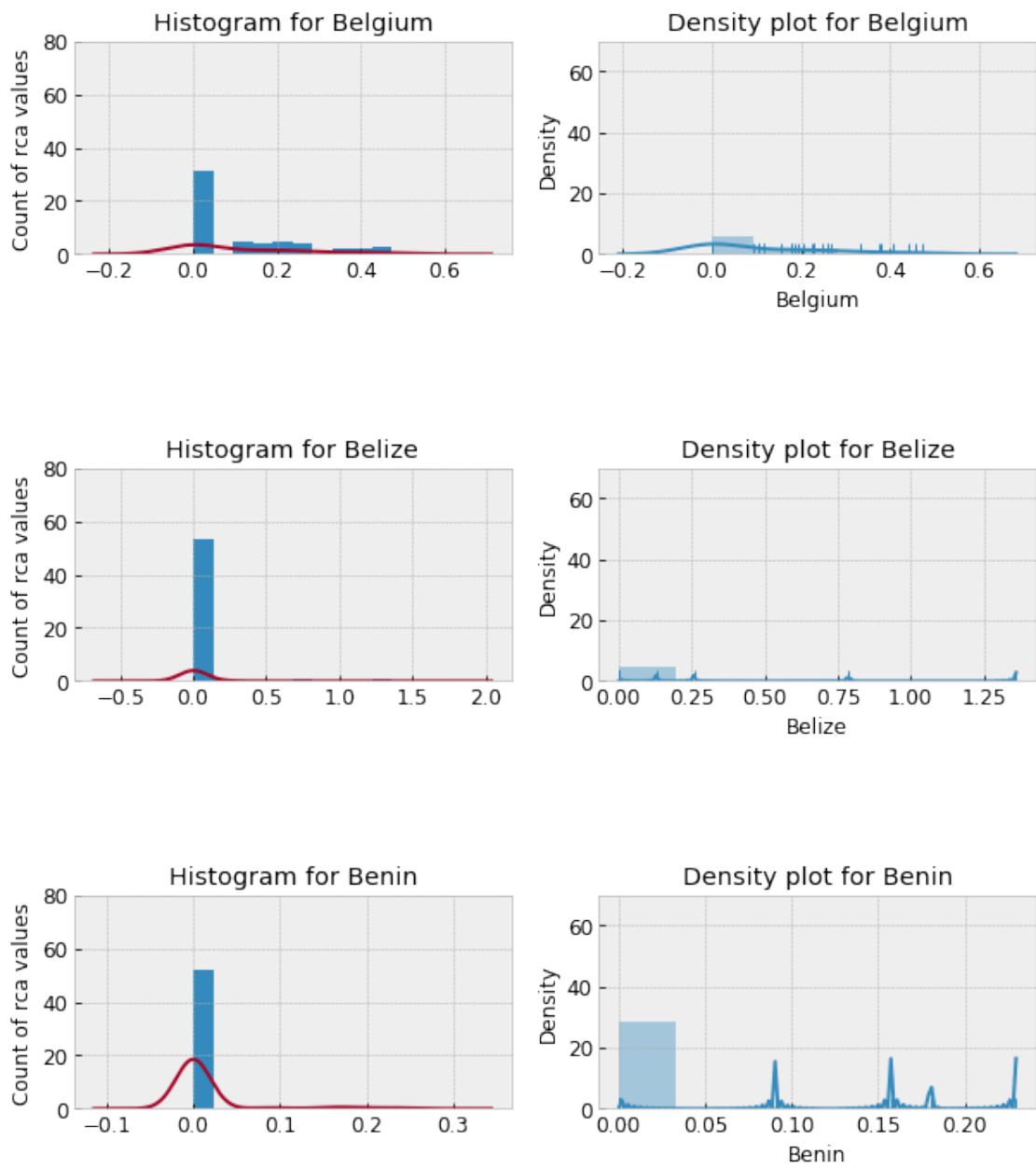
```

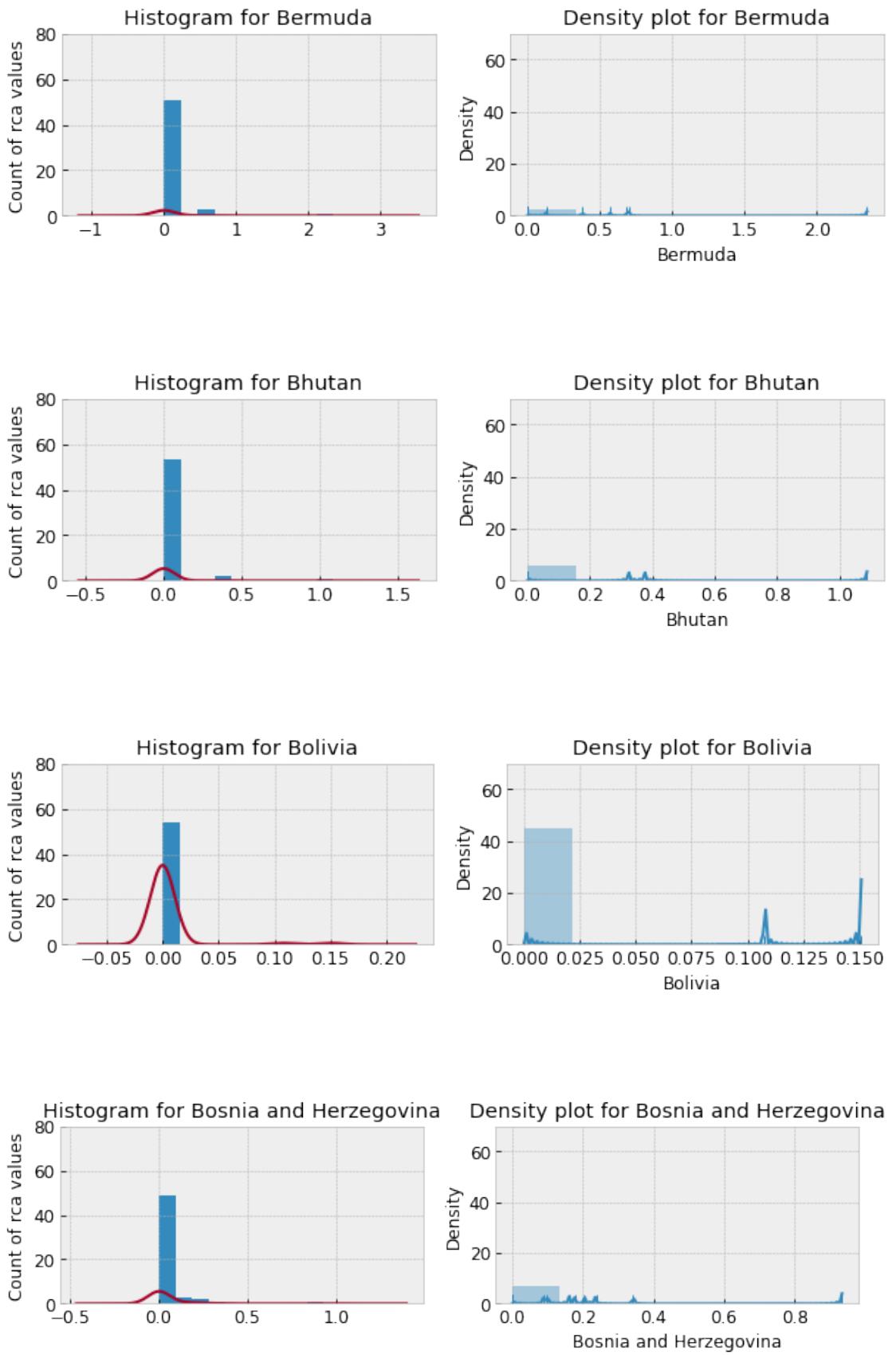


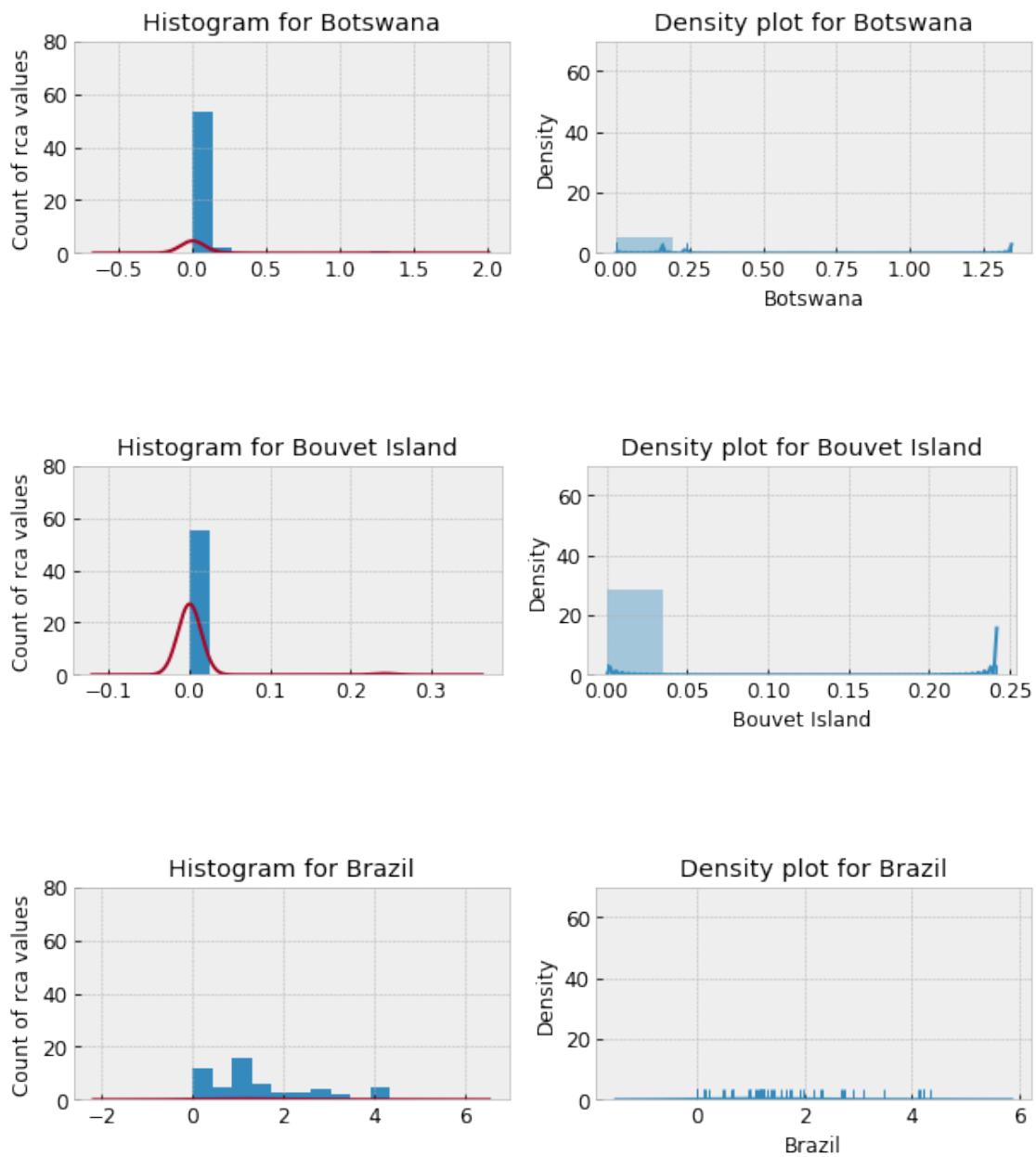


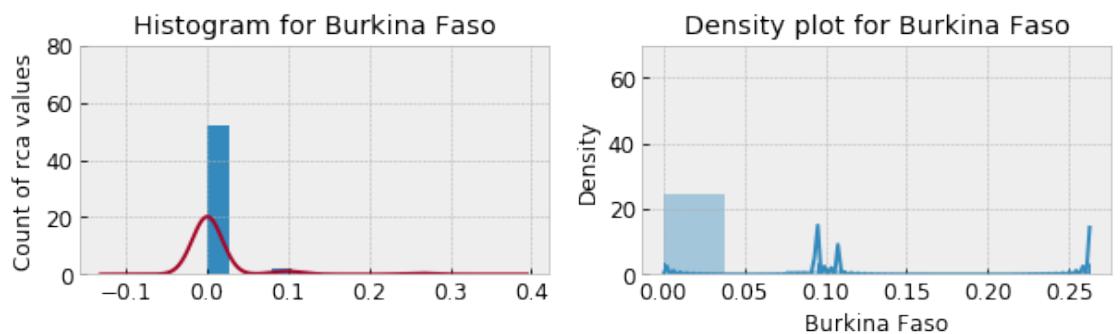
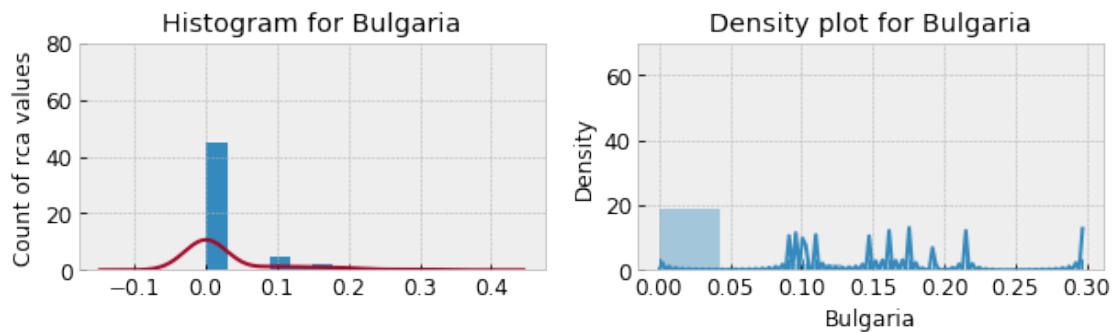
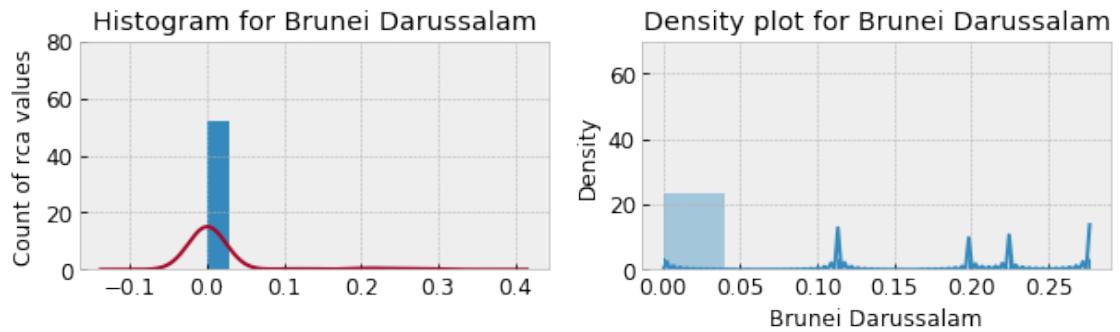
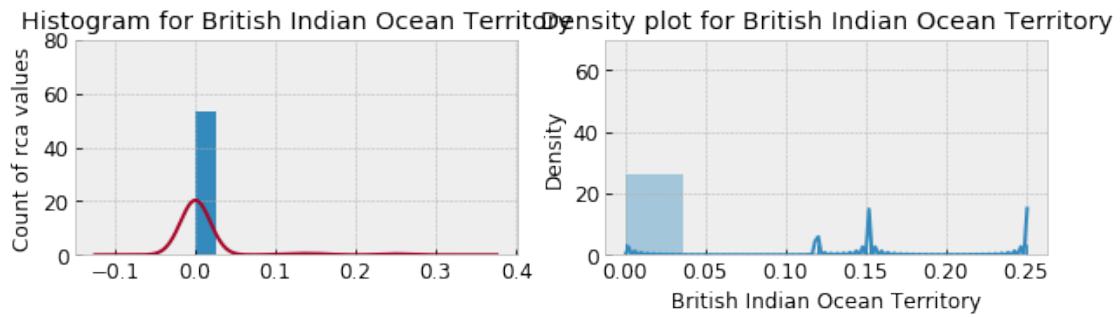


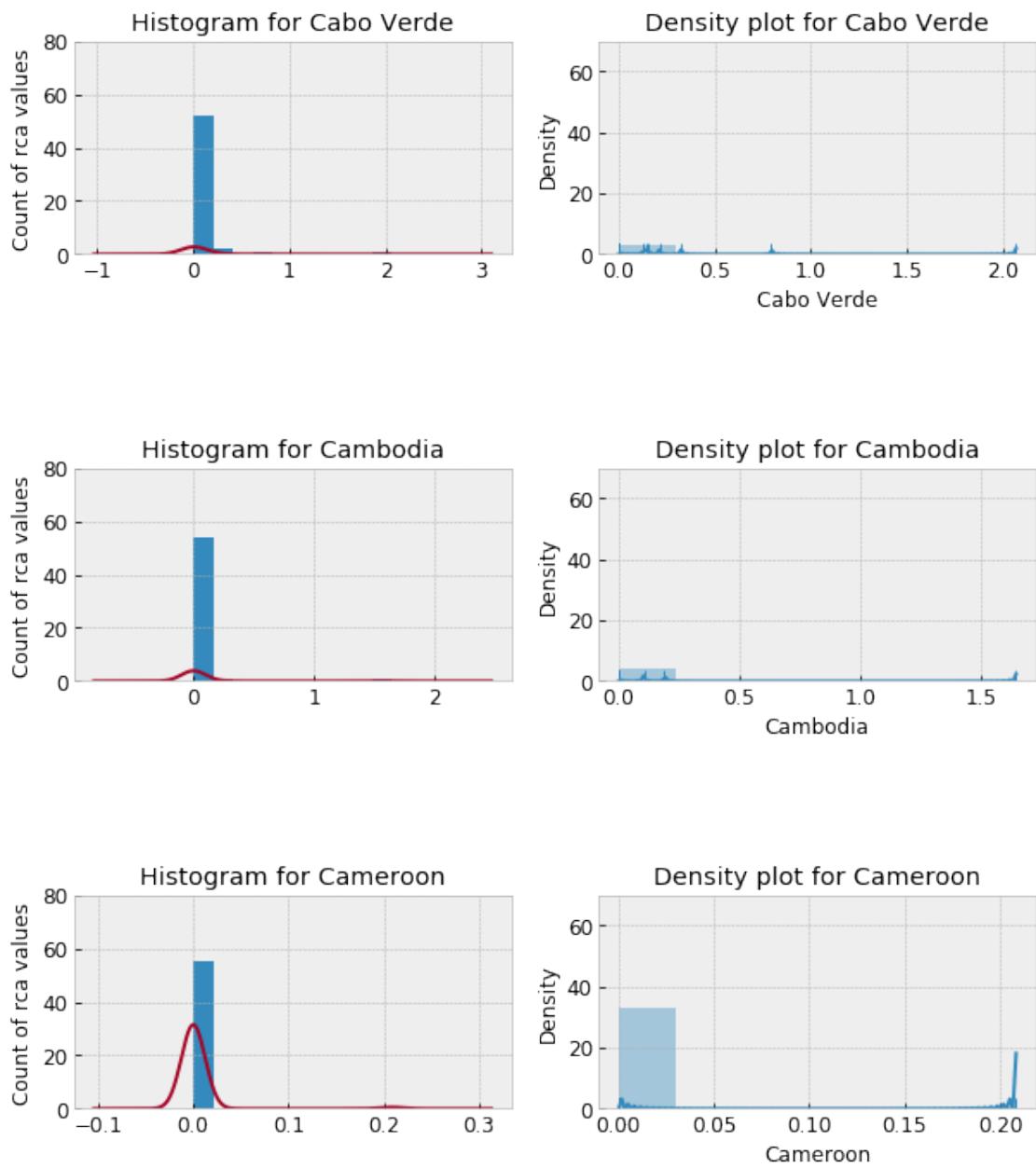


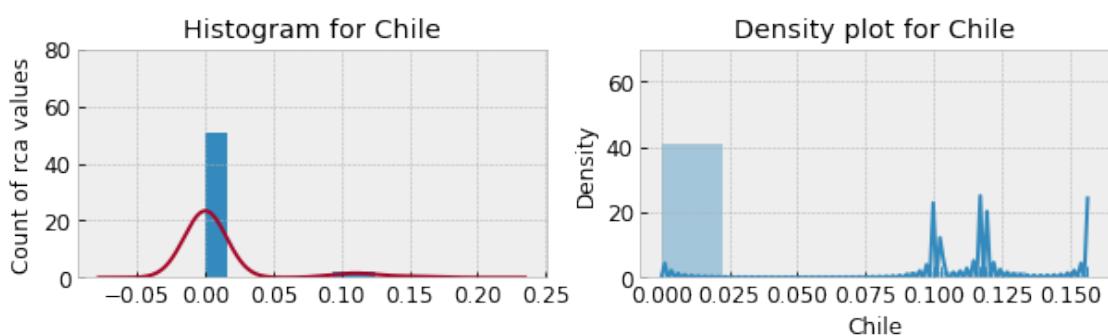
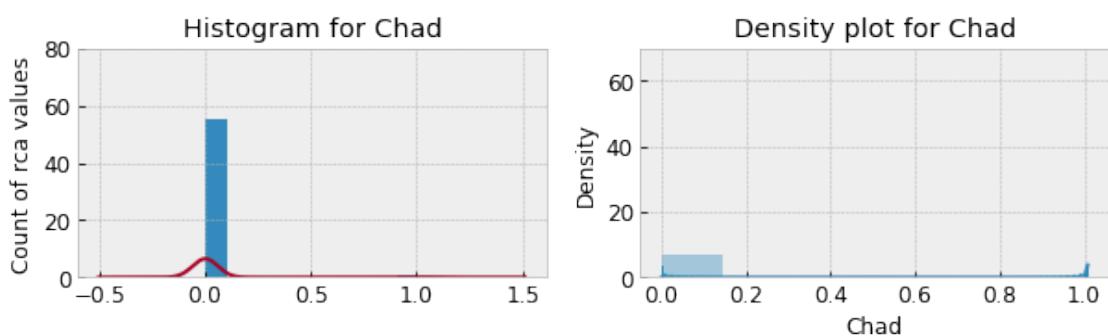
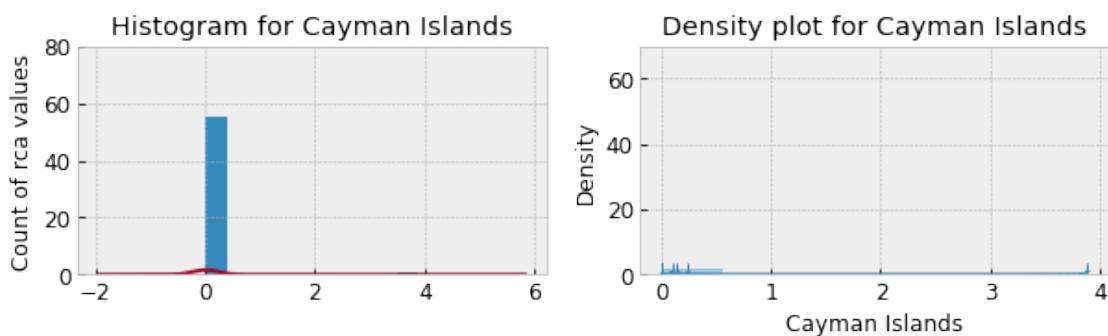
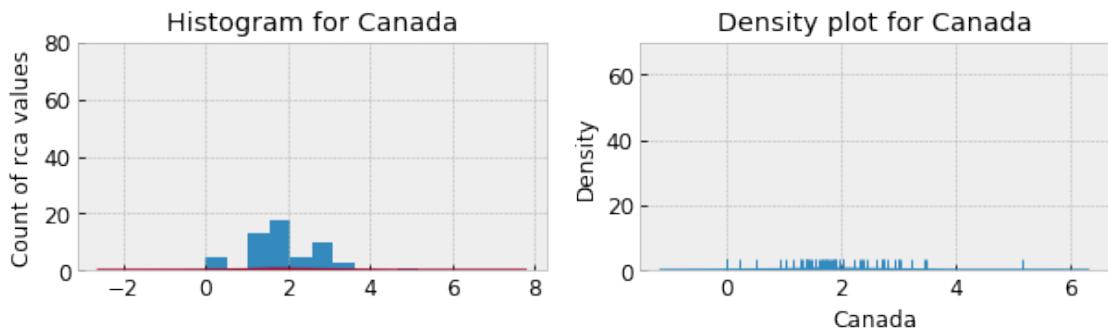


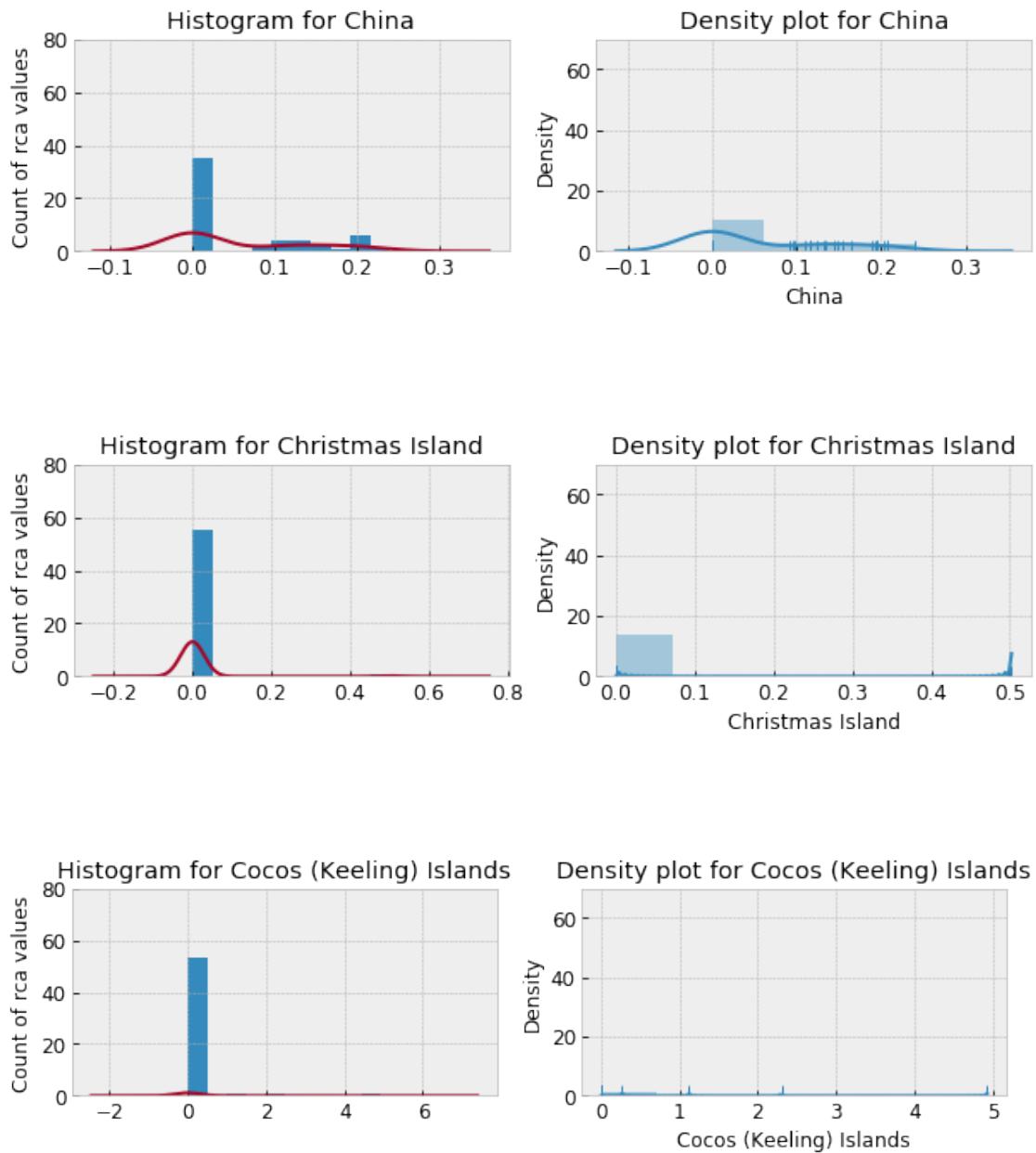


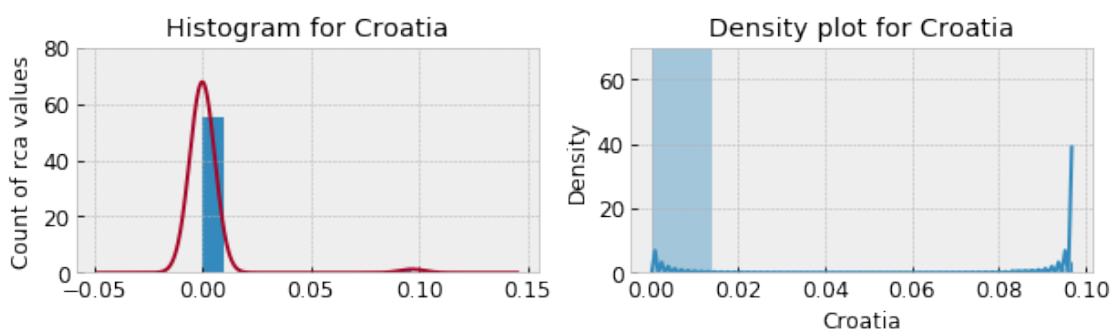
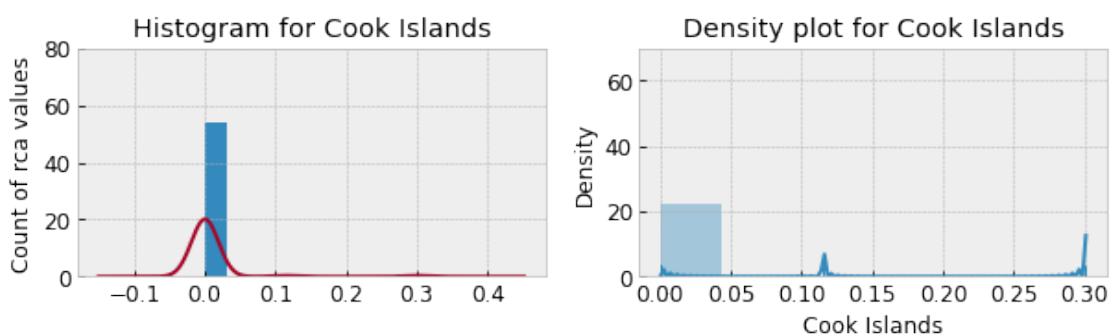
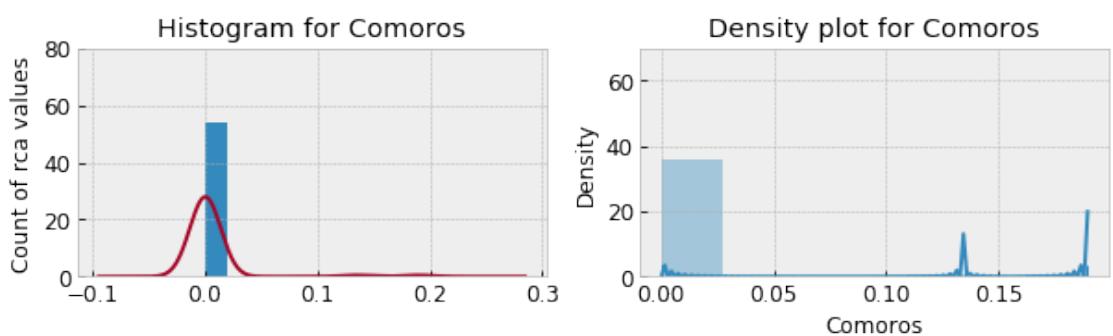
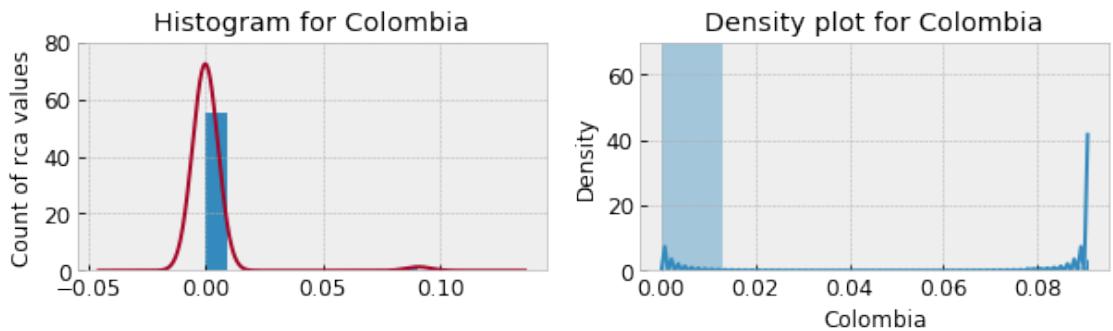


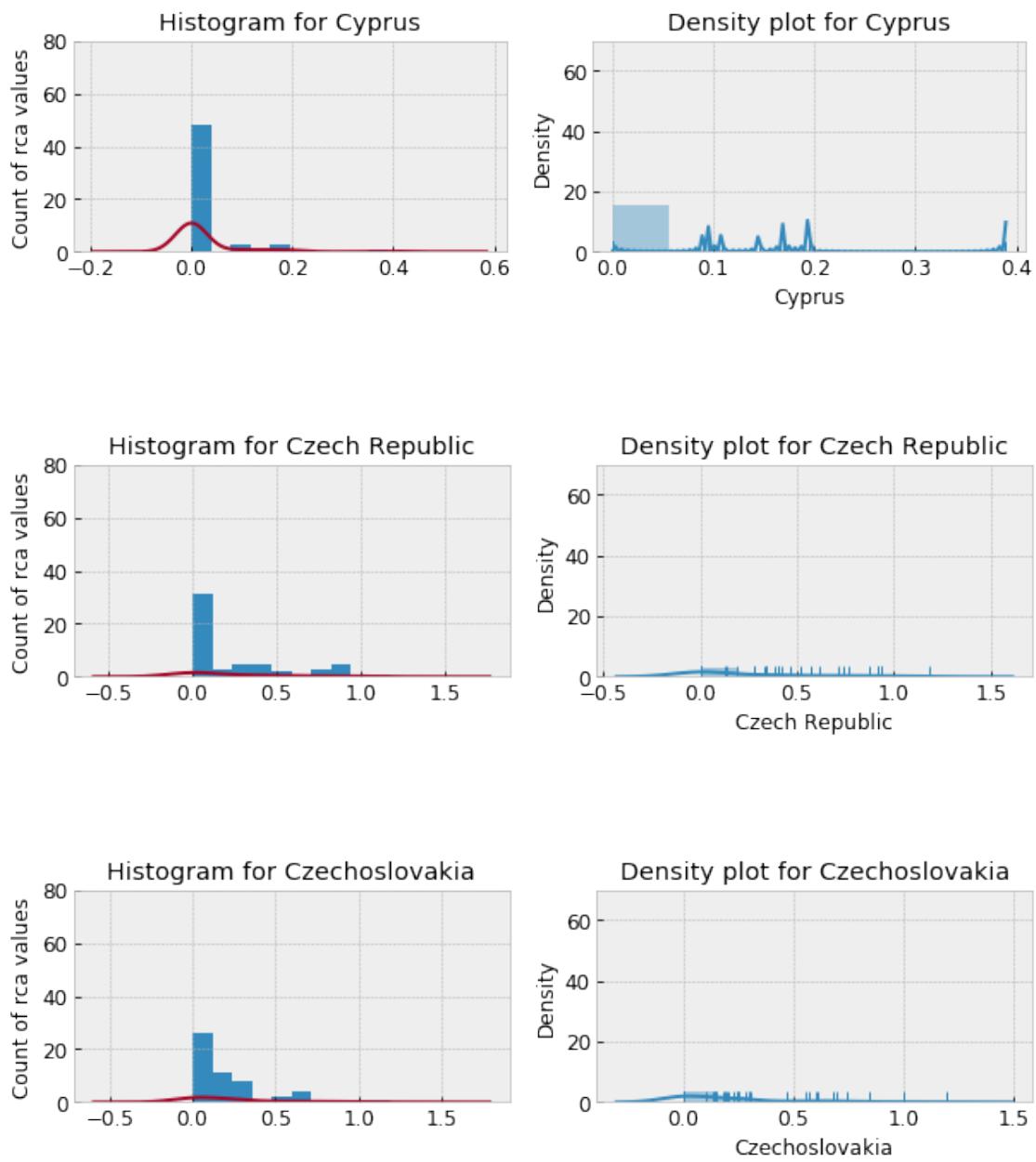


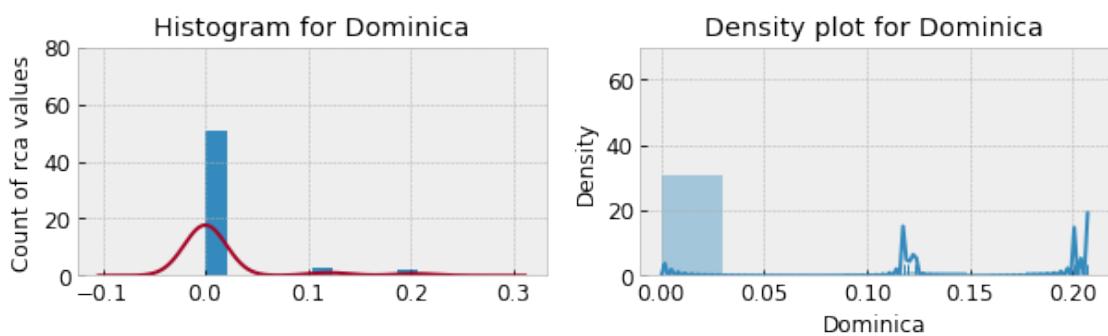
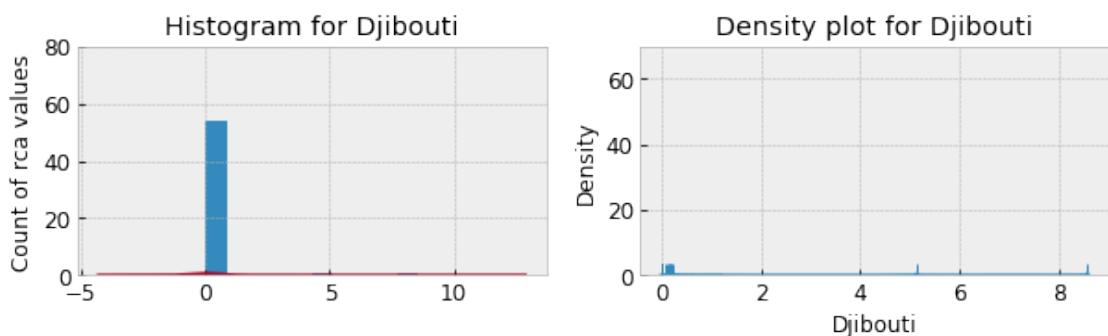
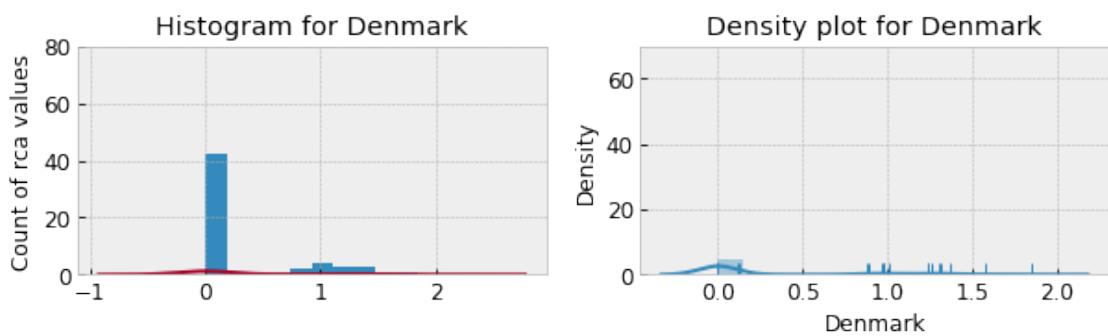
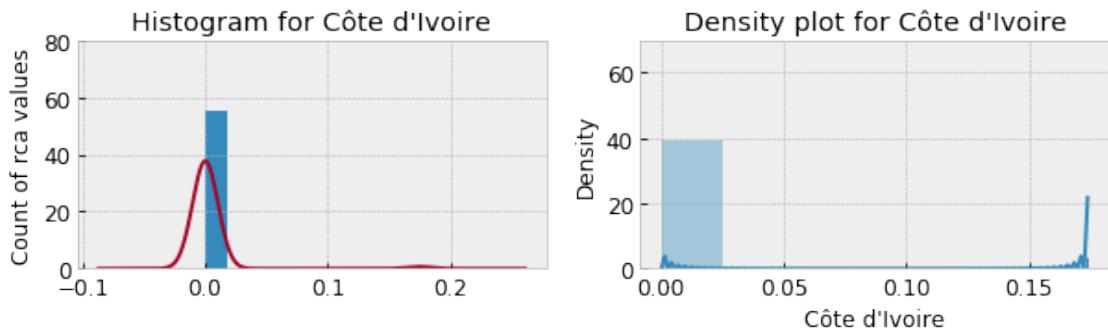


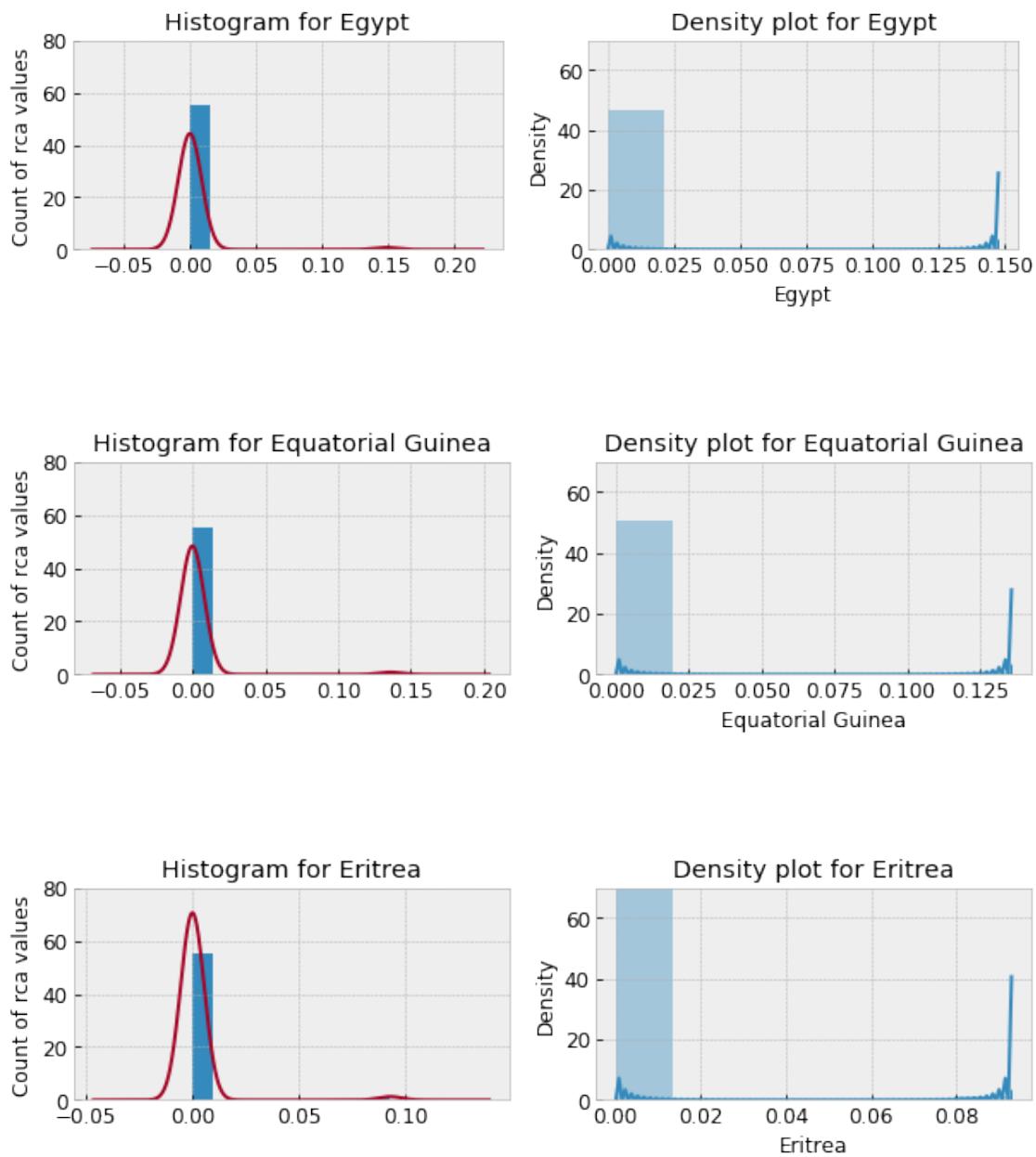


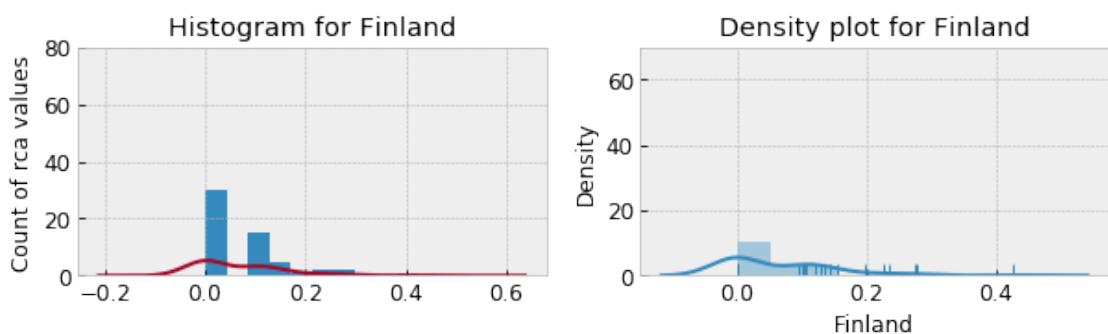
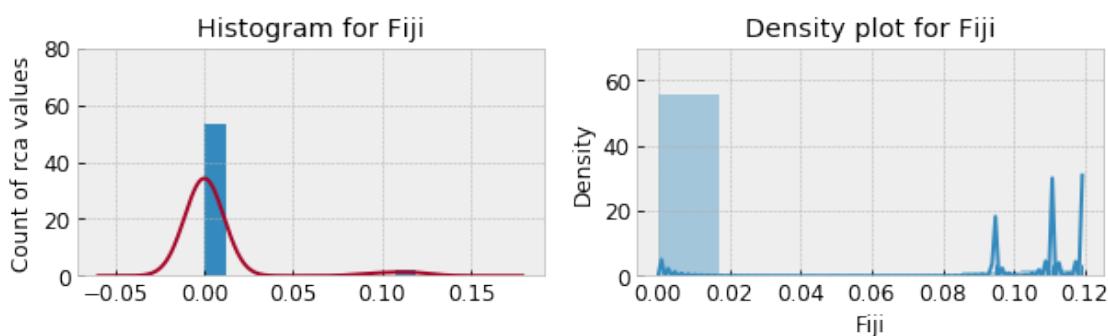
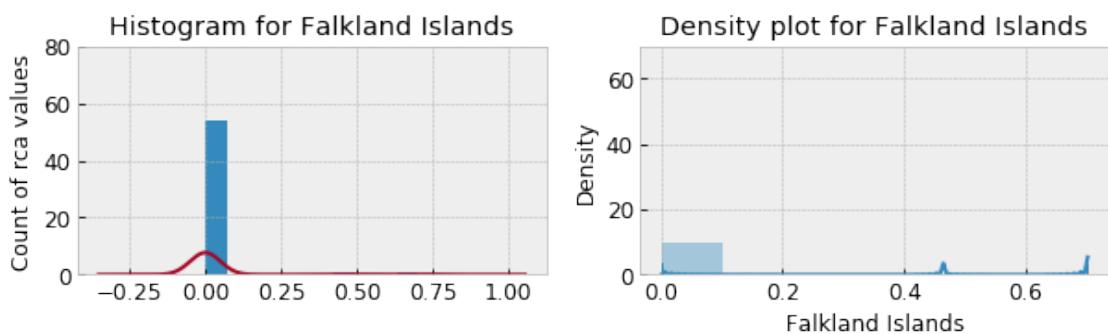
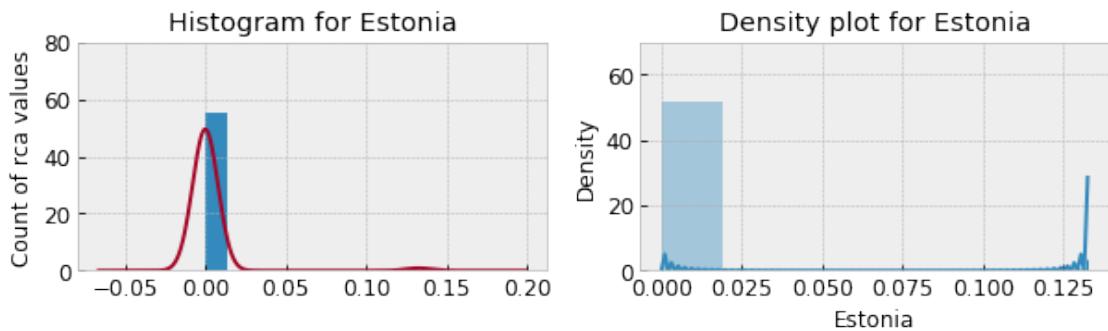




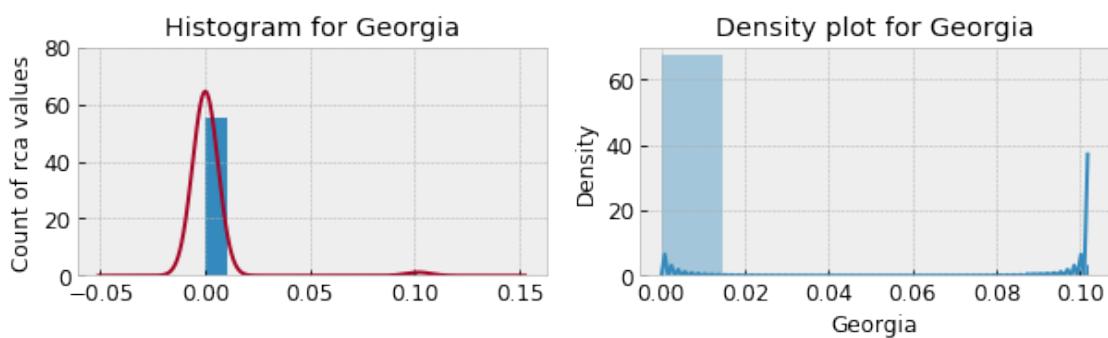
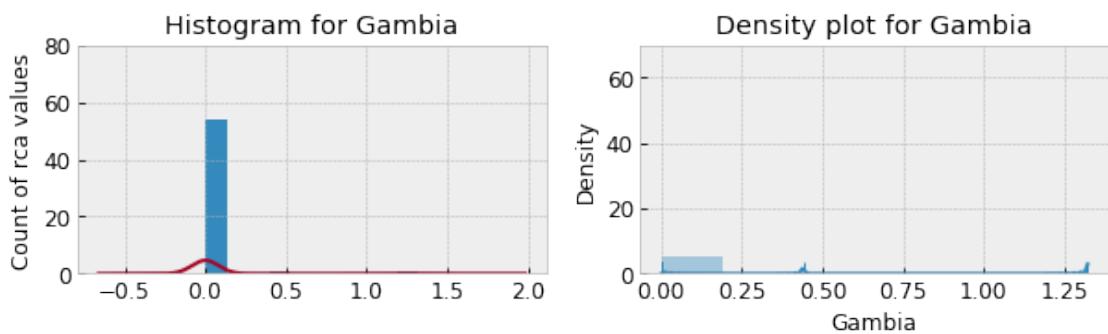
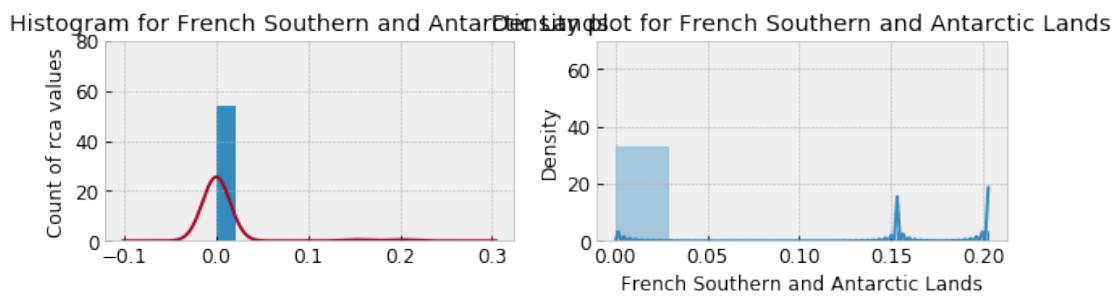
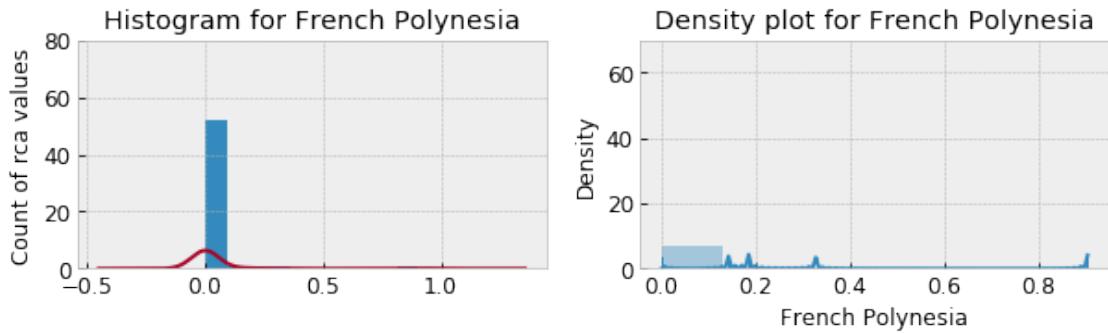


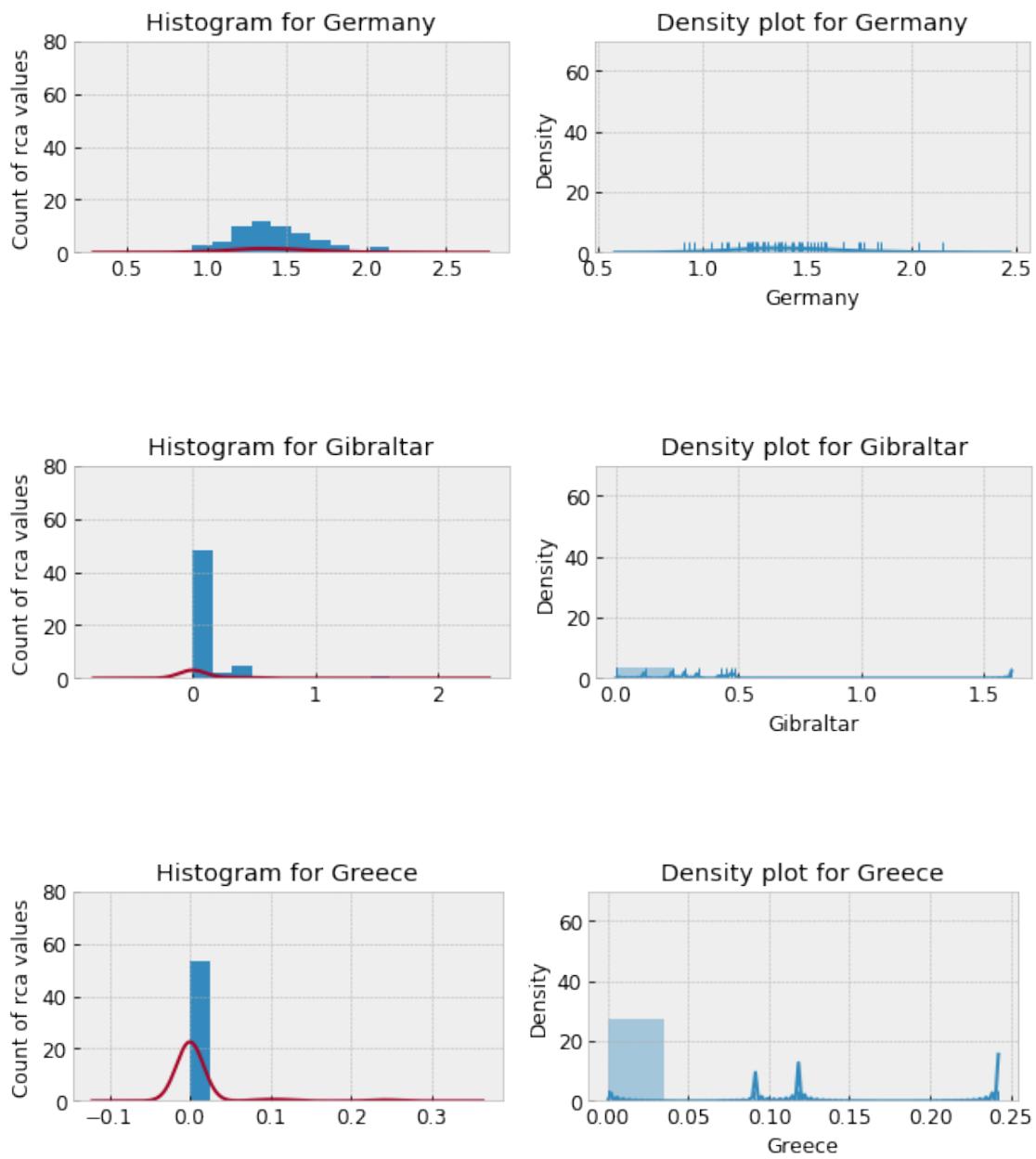


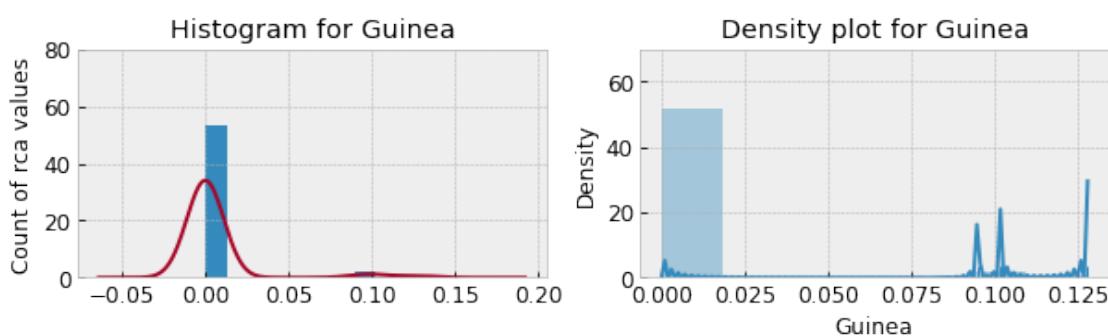
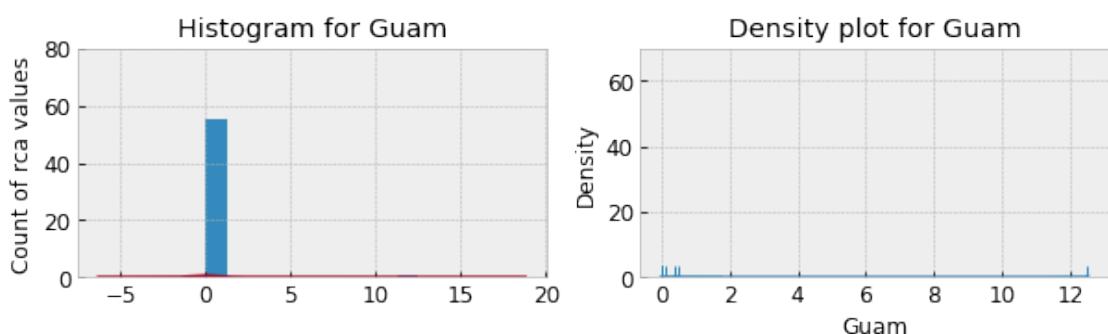
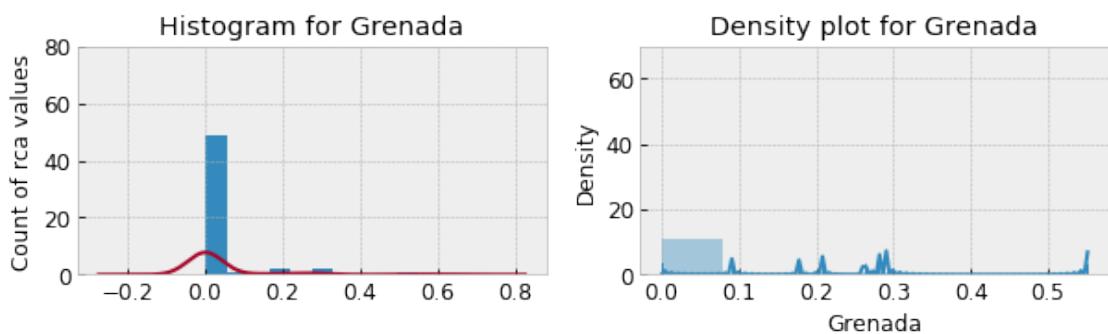
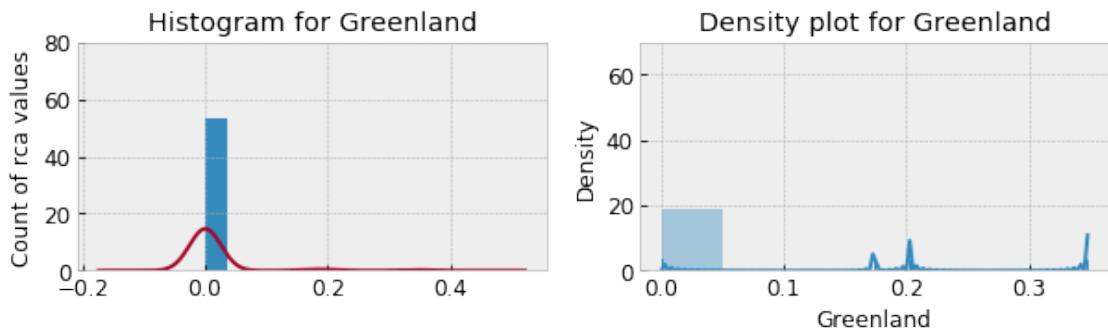


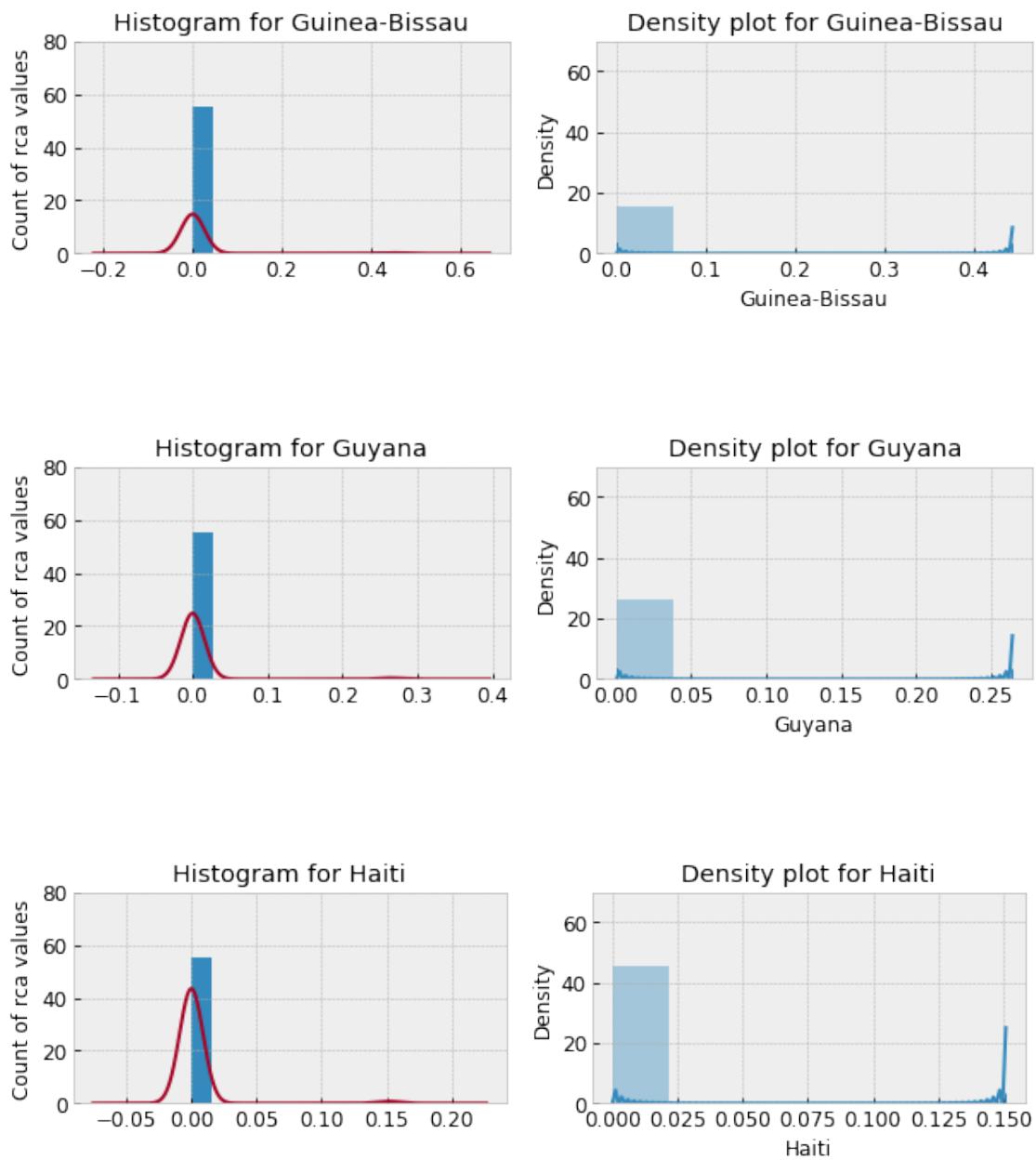


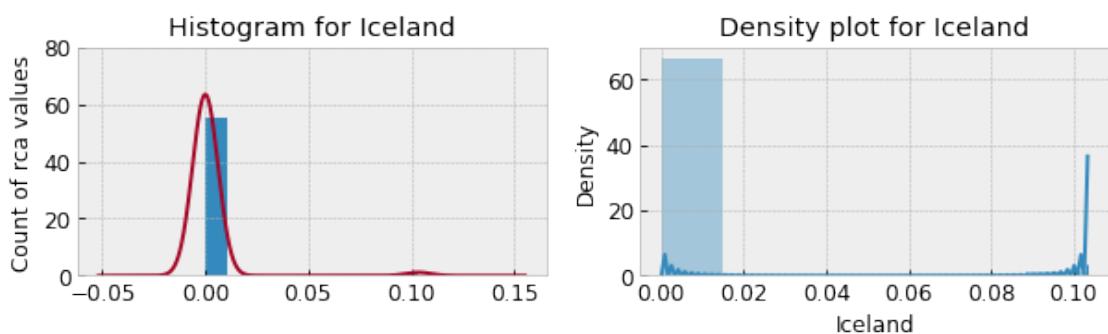
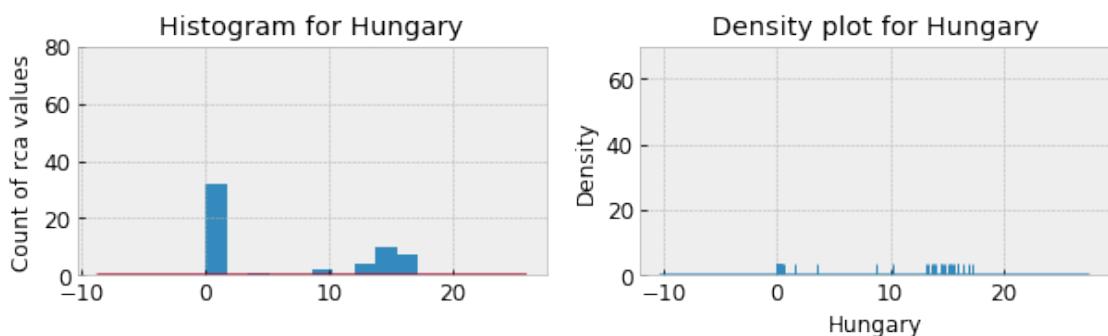
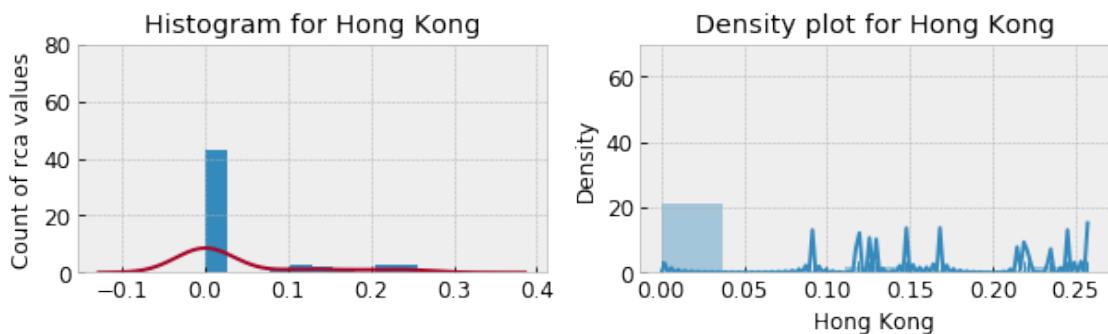
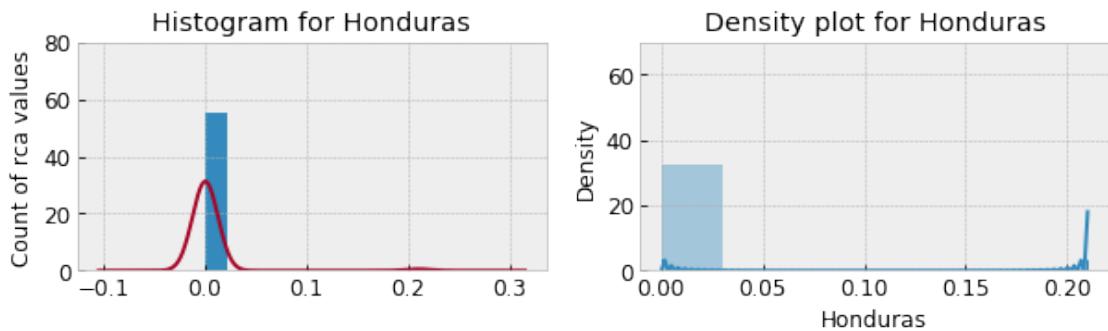


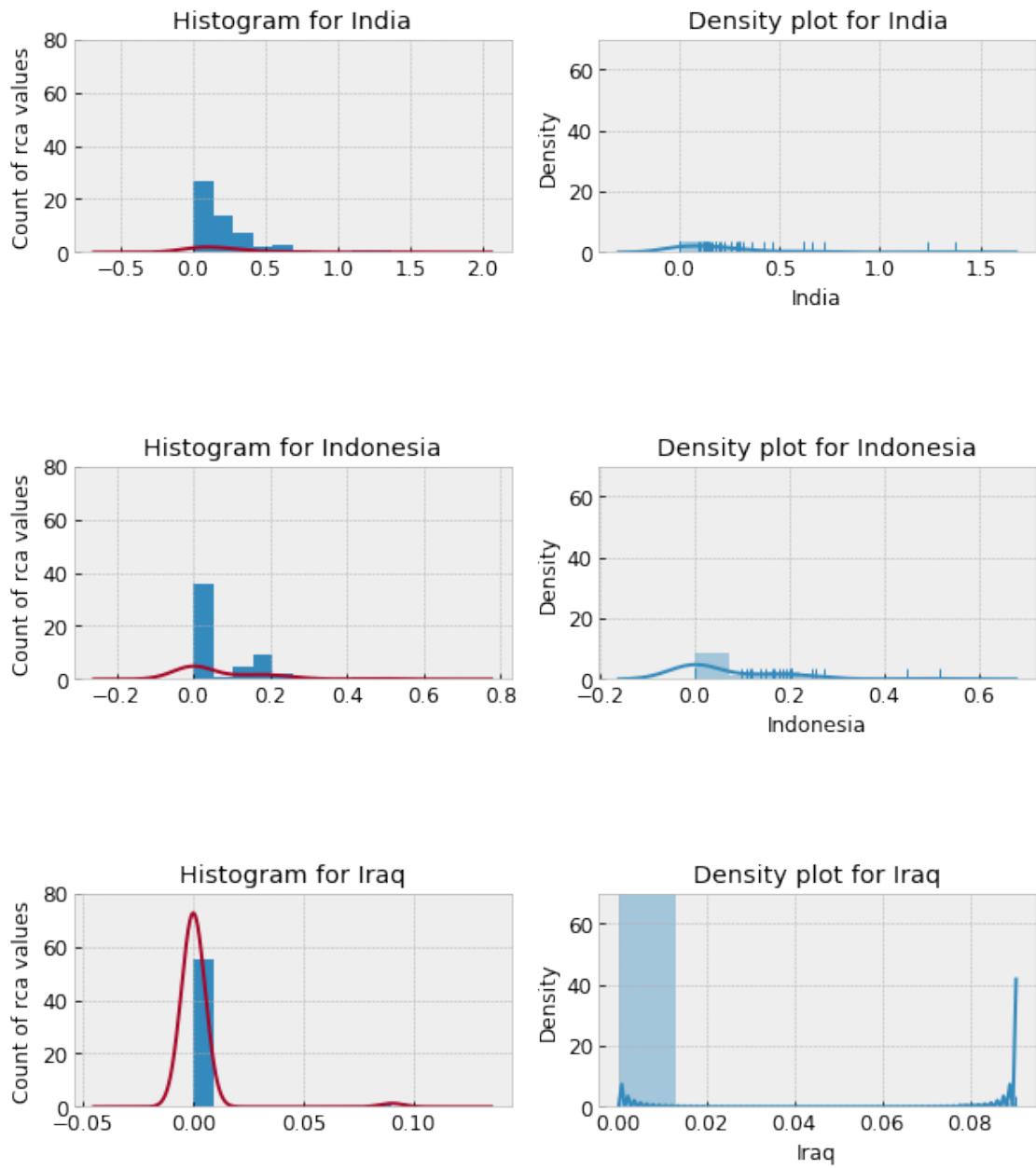


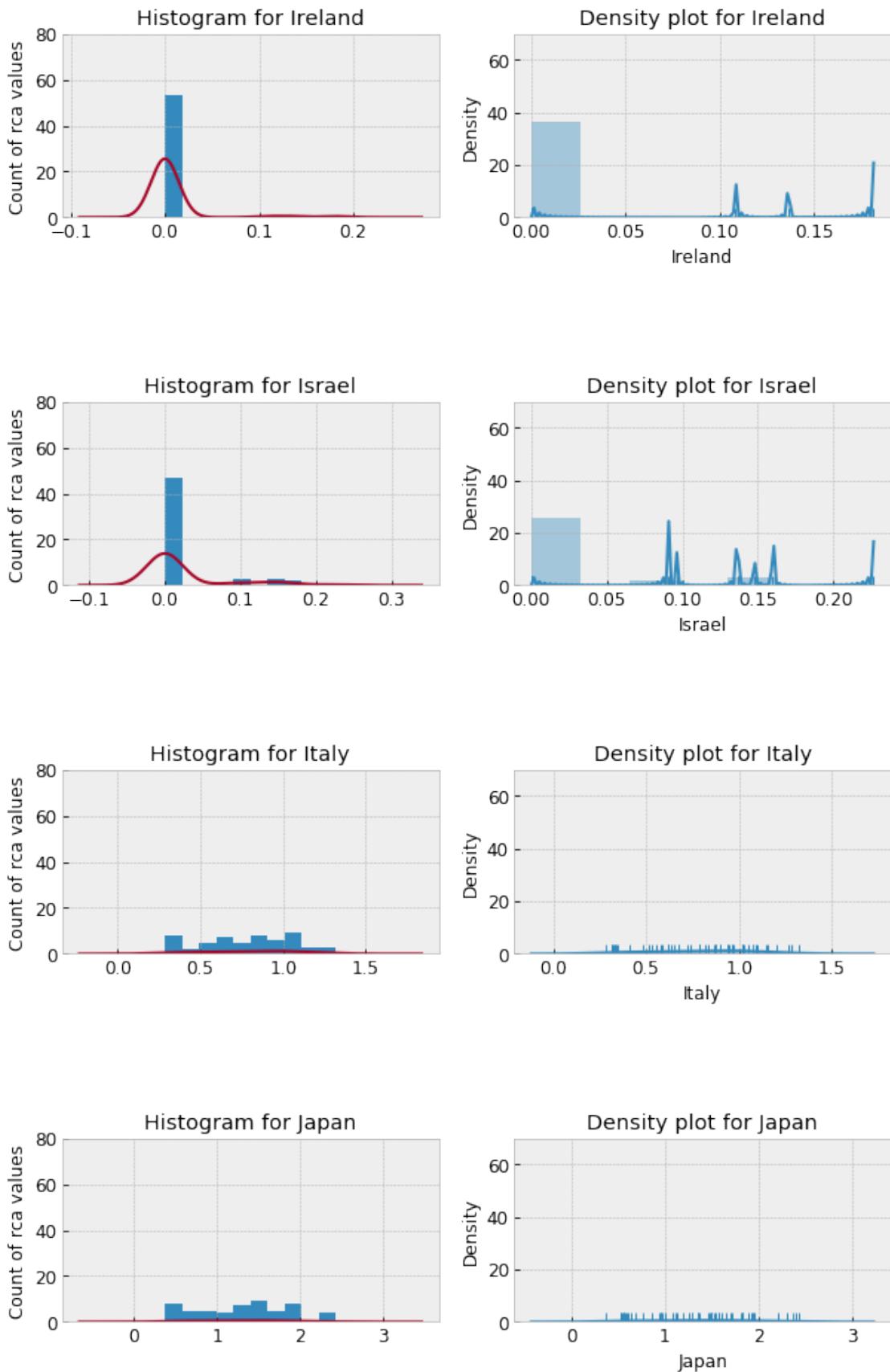


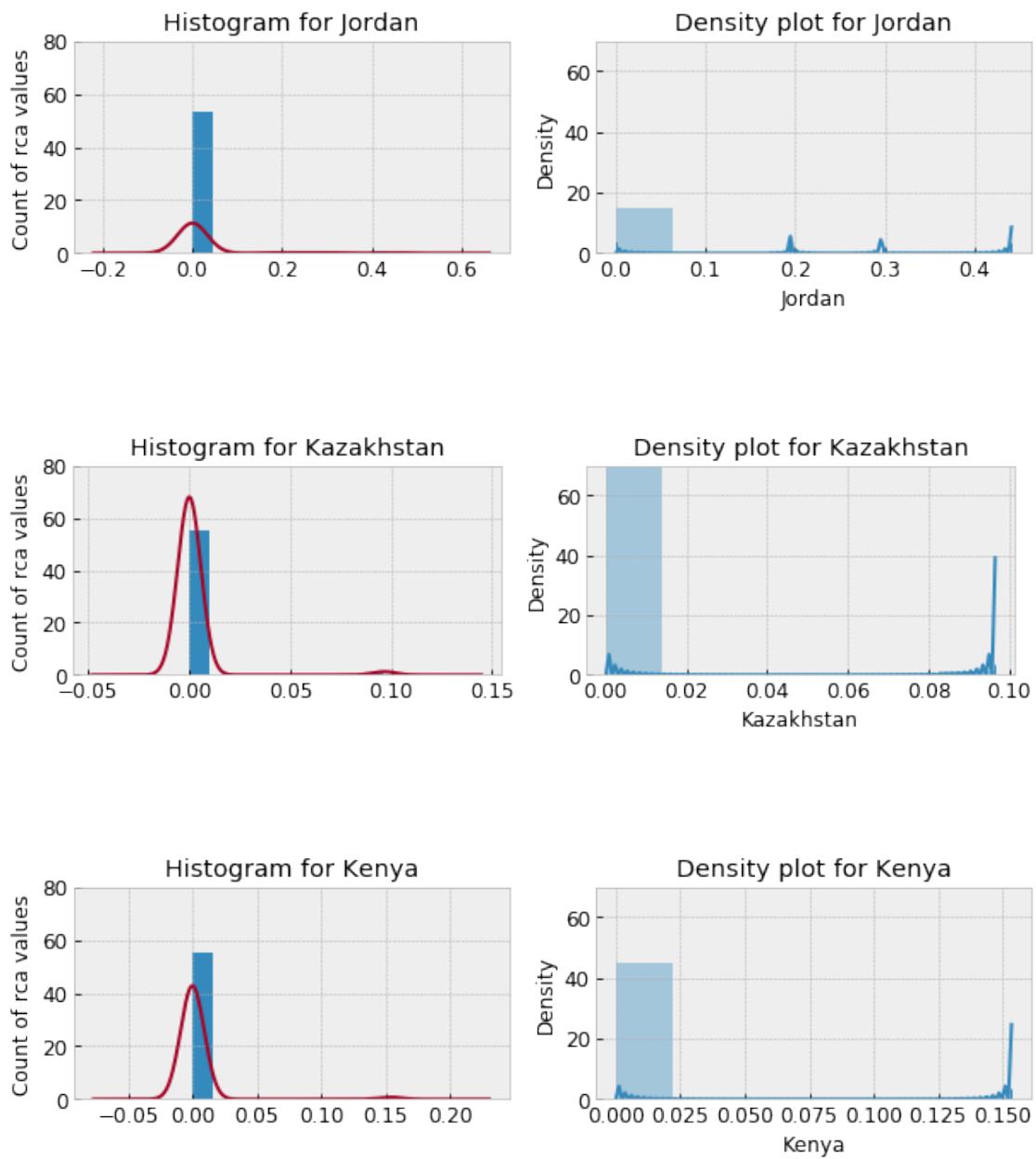


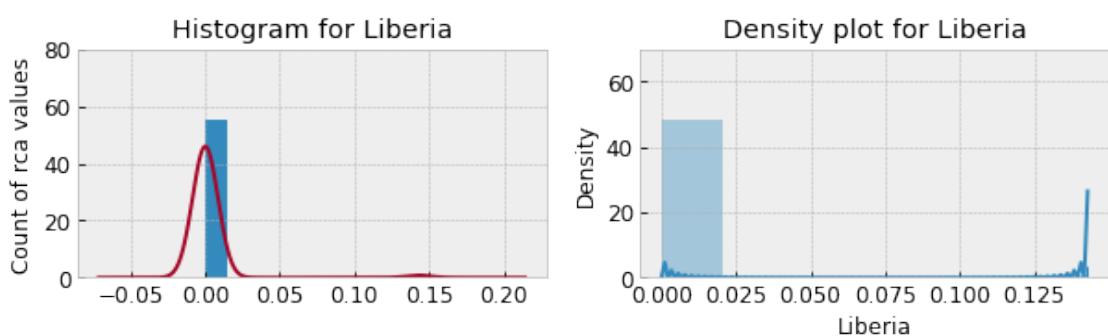
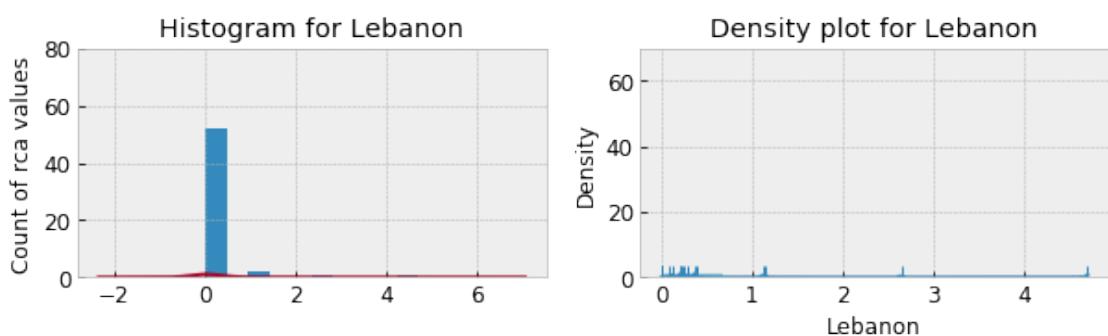
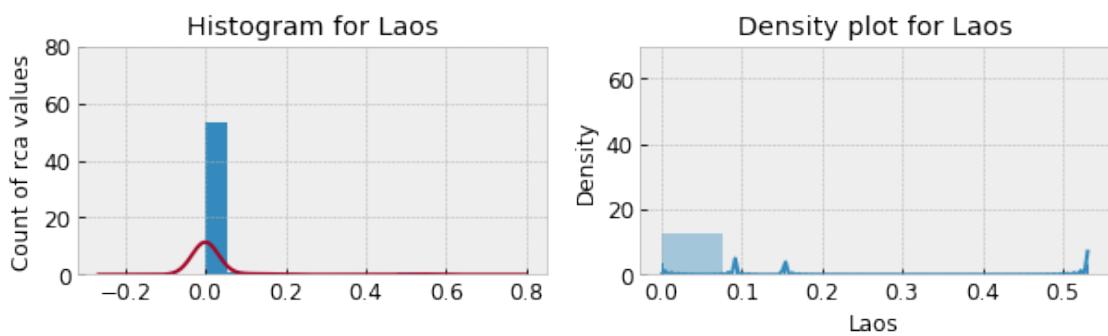
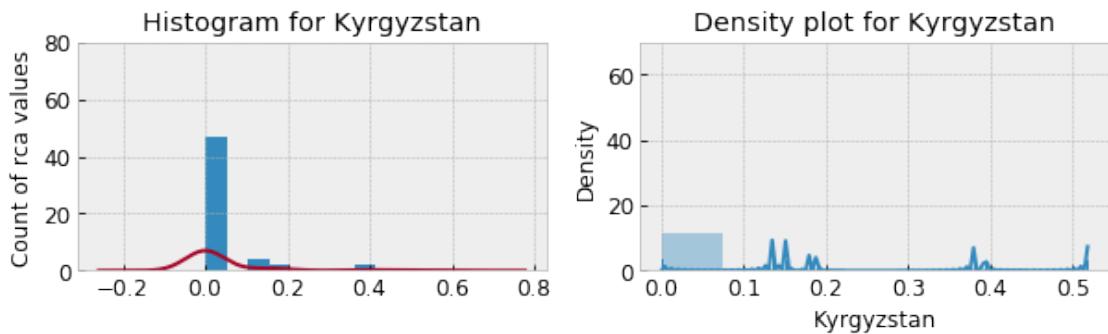


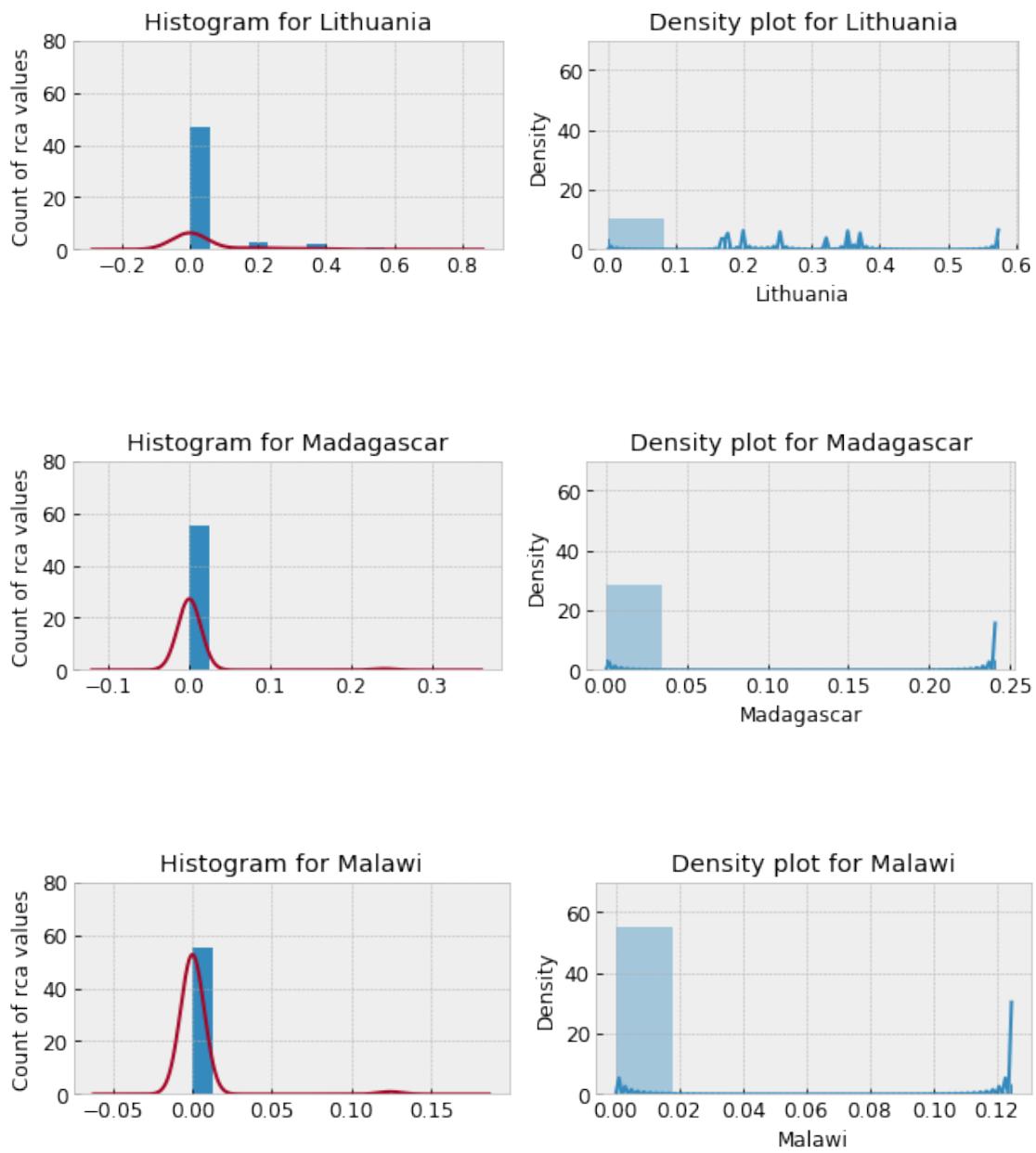


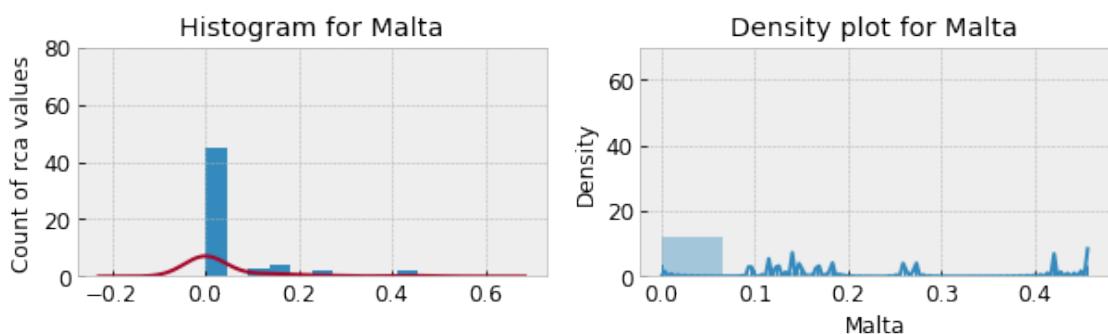
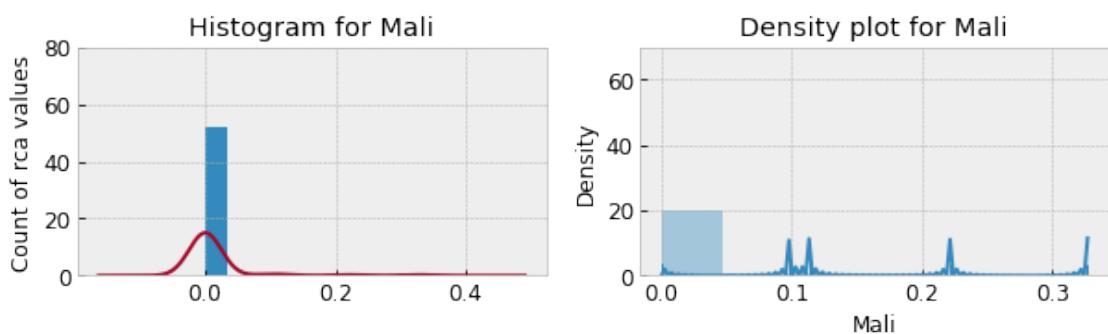
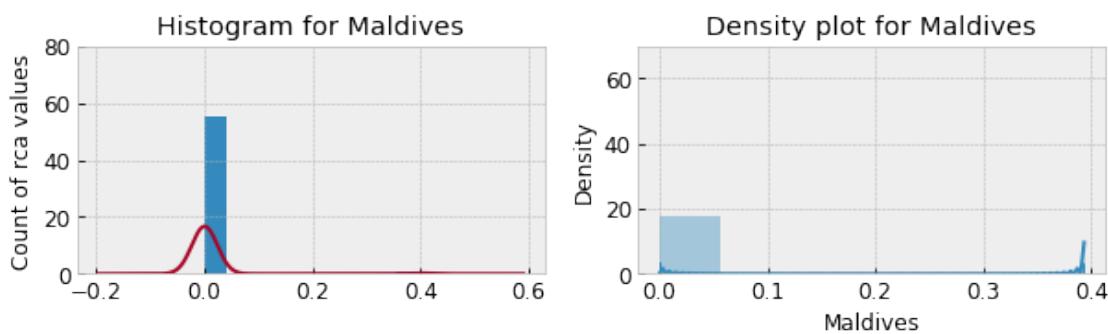
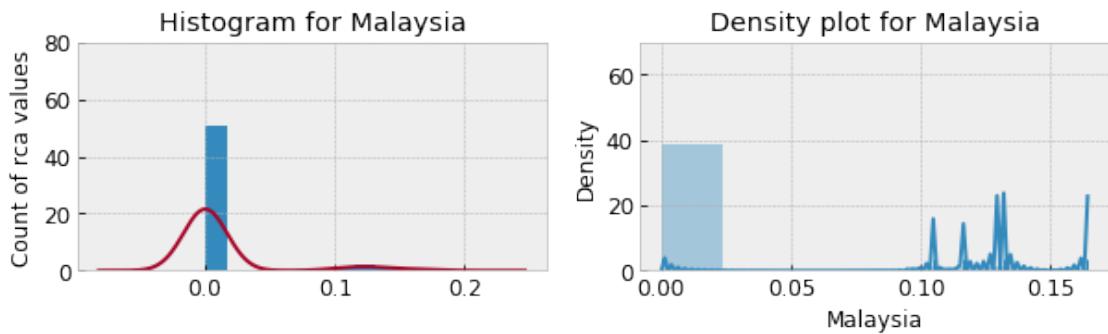


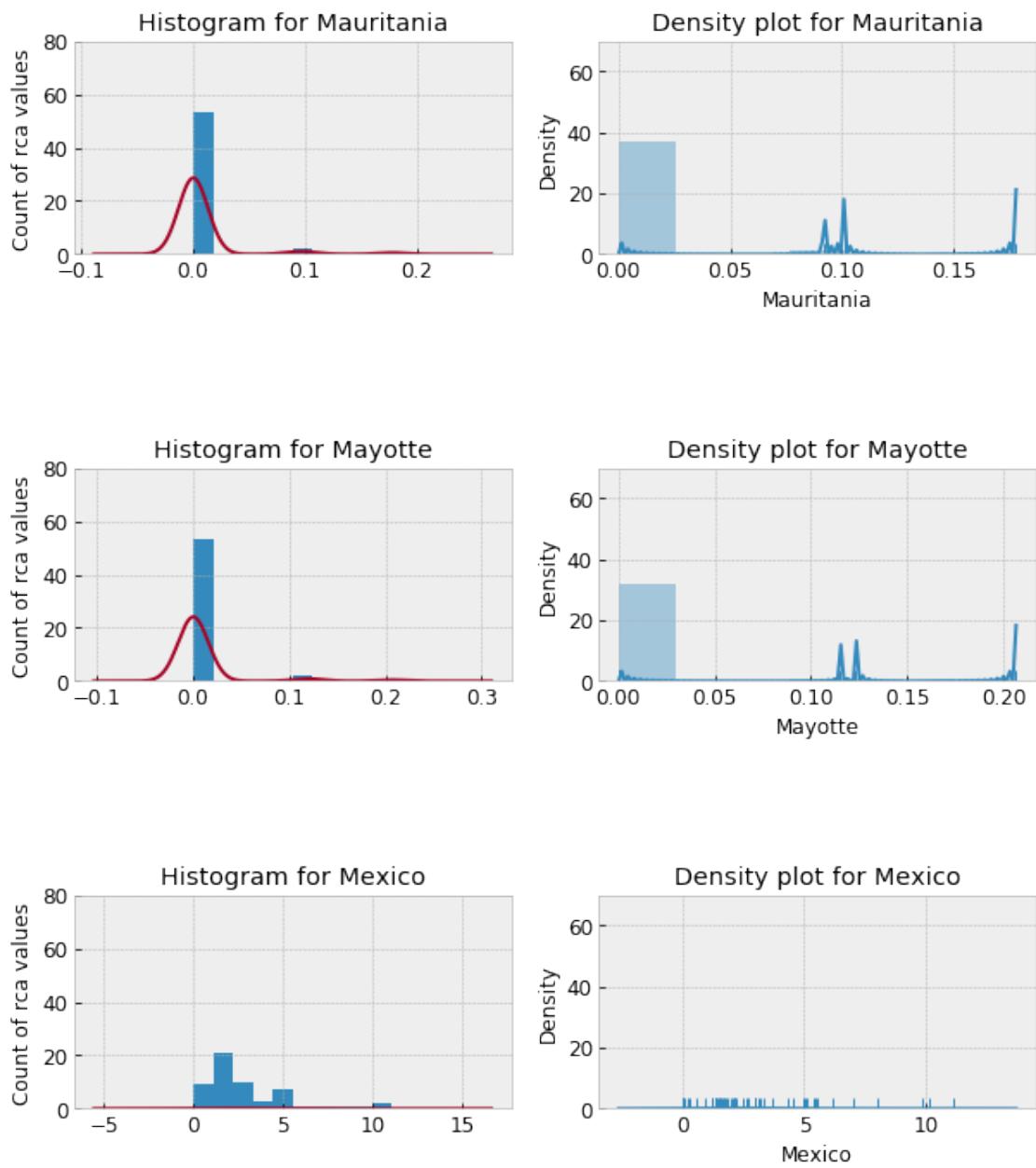


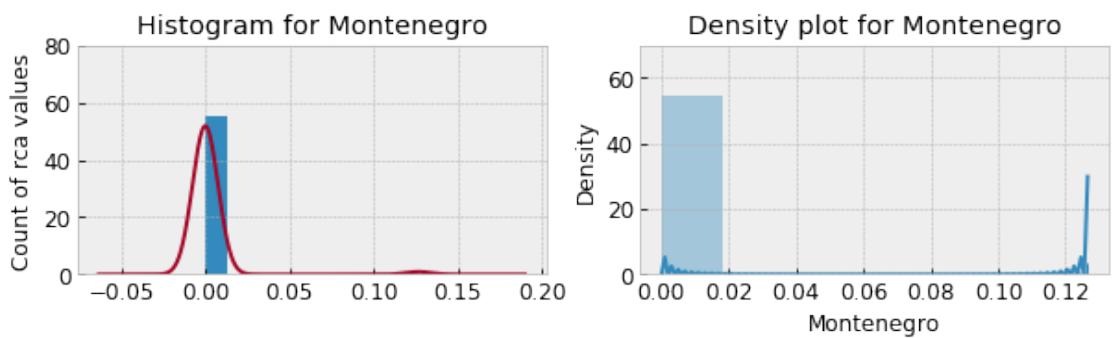
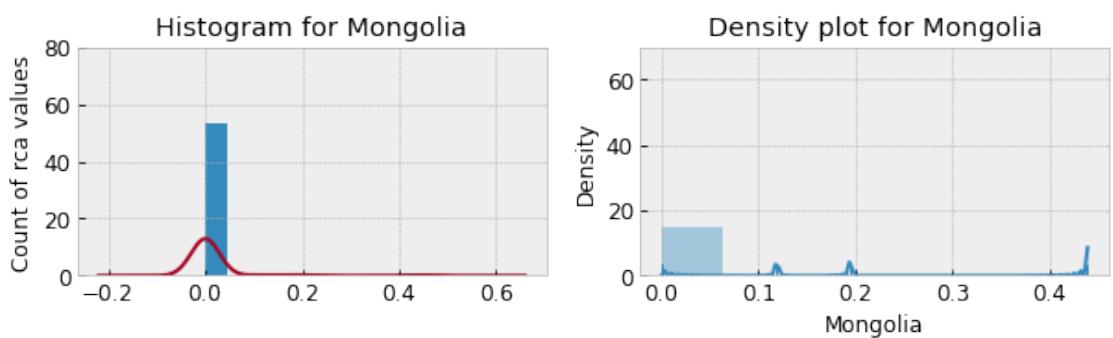
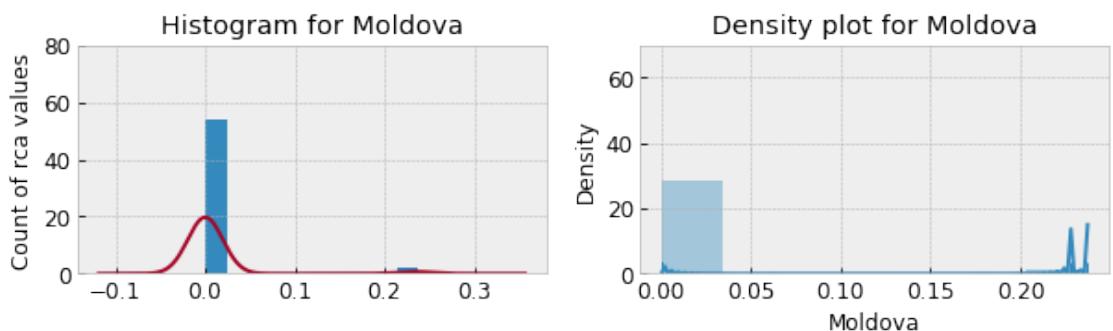
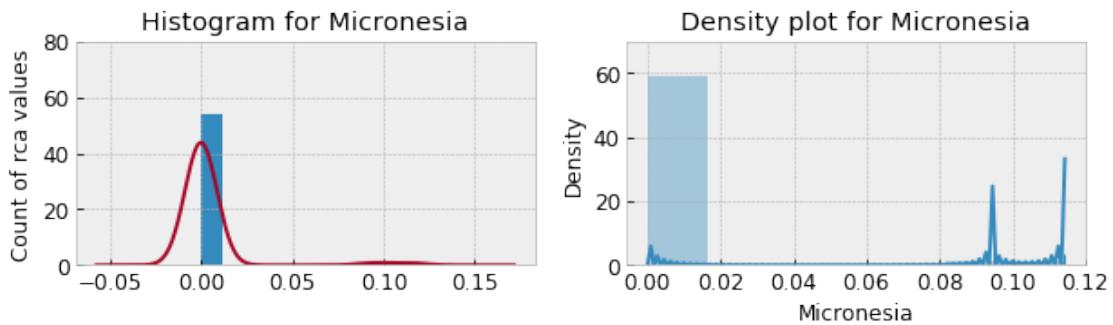


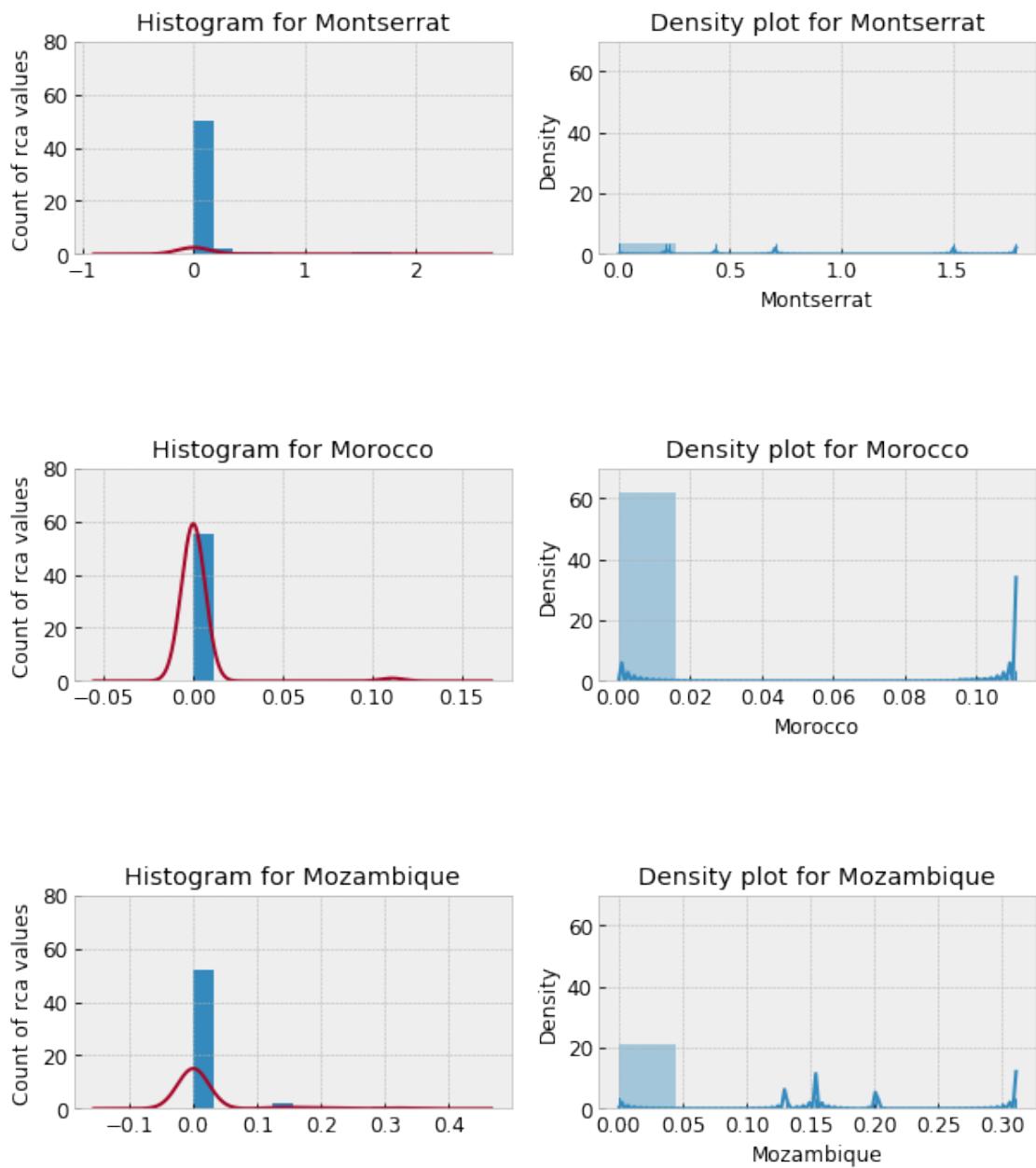


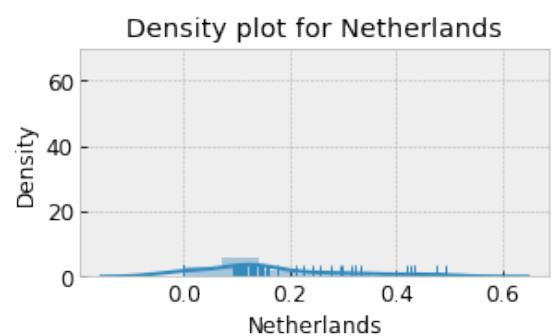
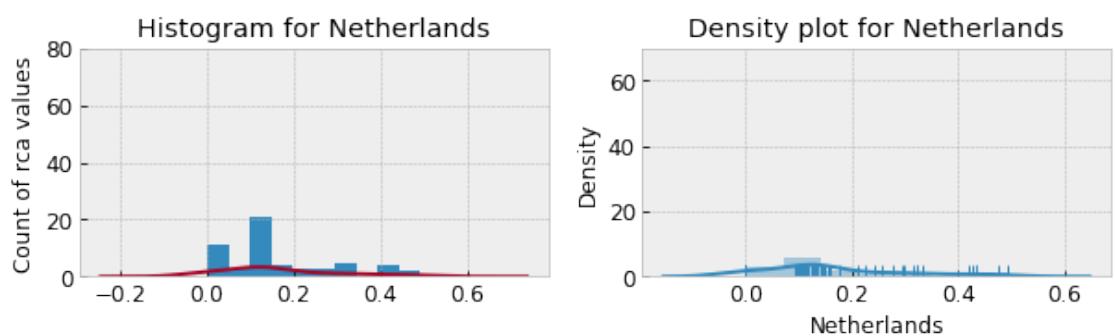
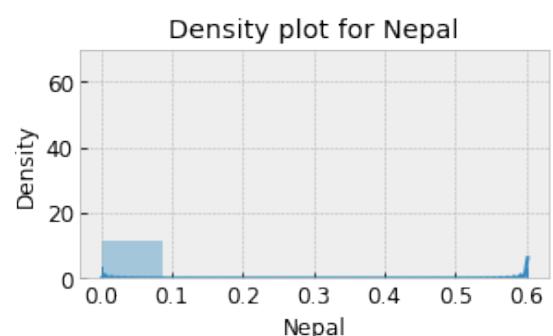
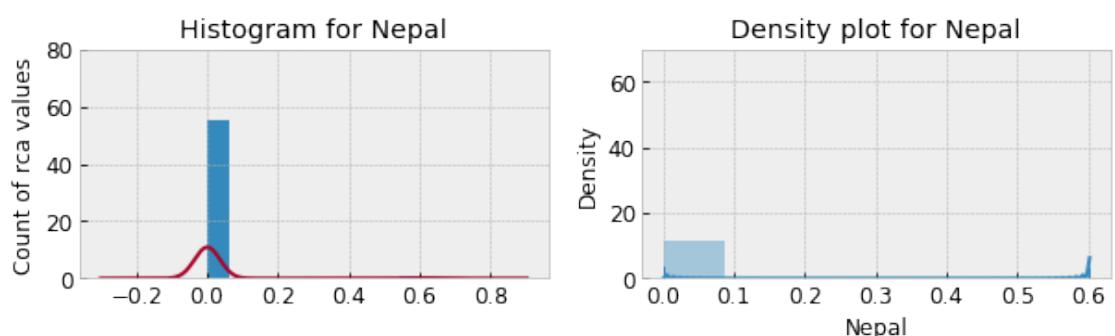
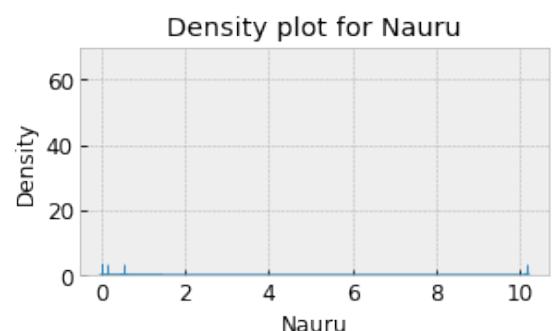
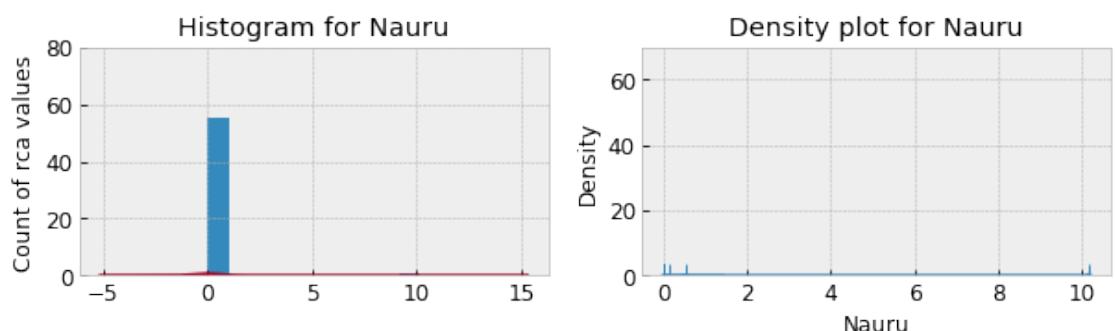
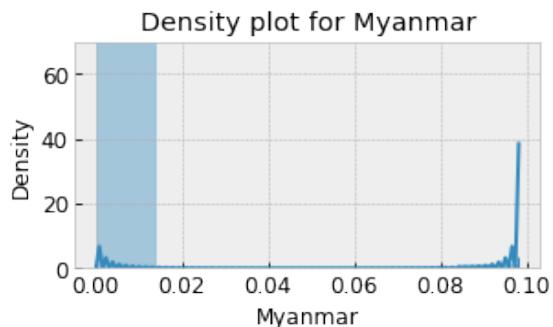
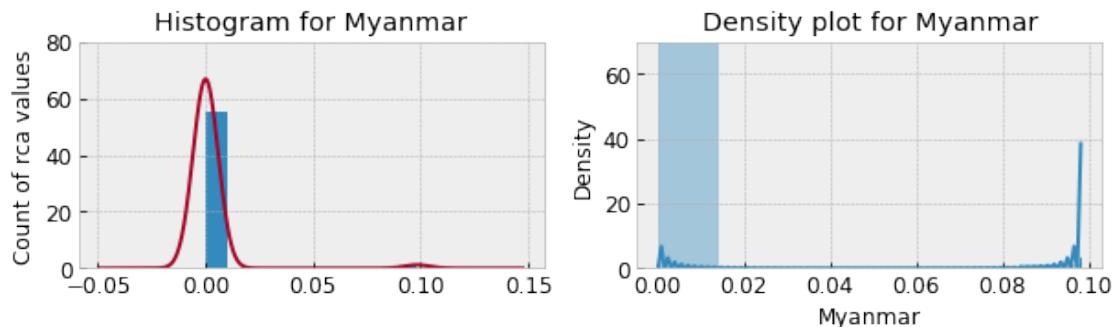


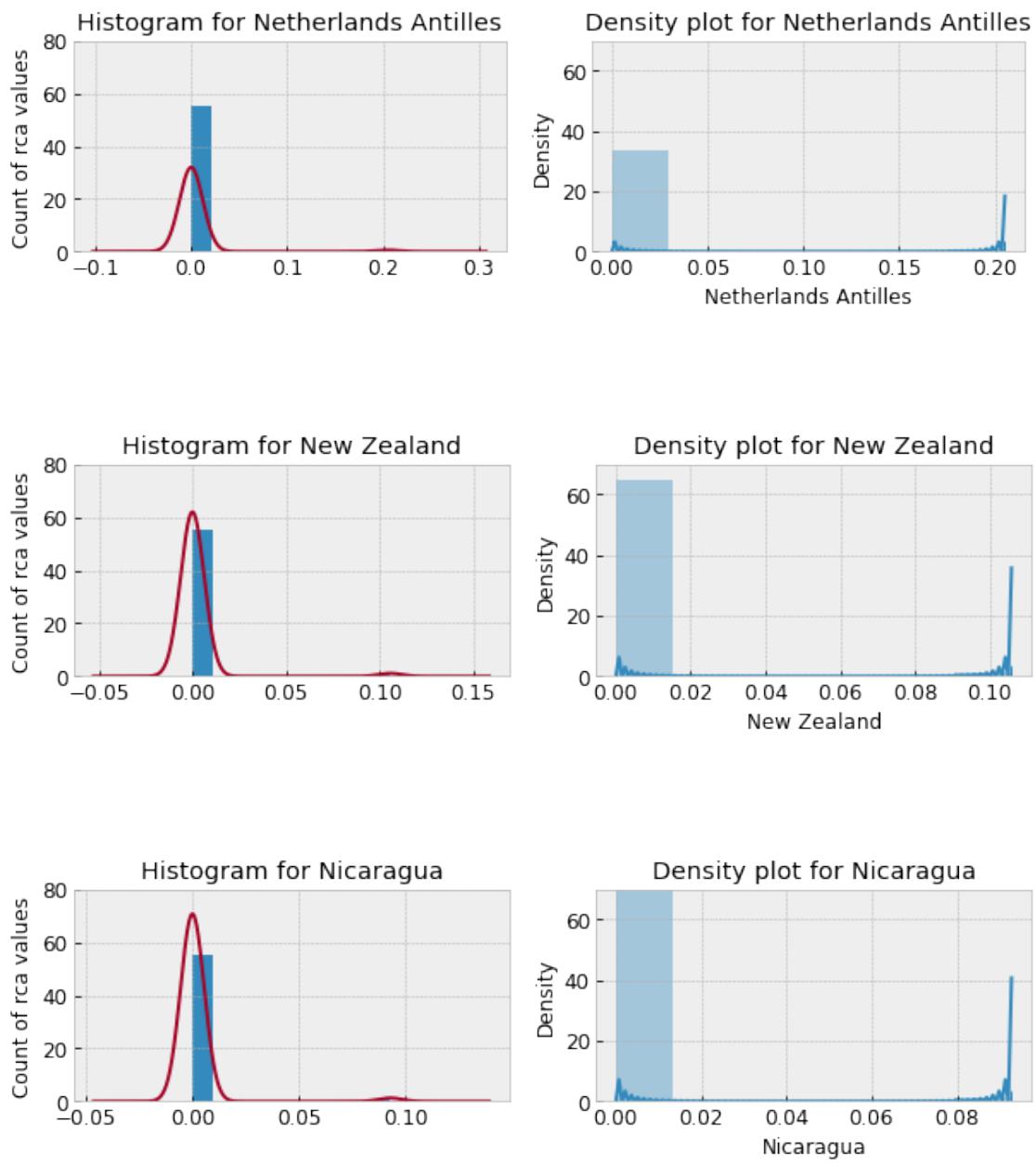


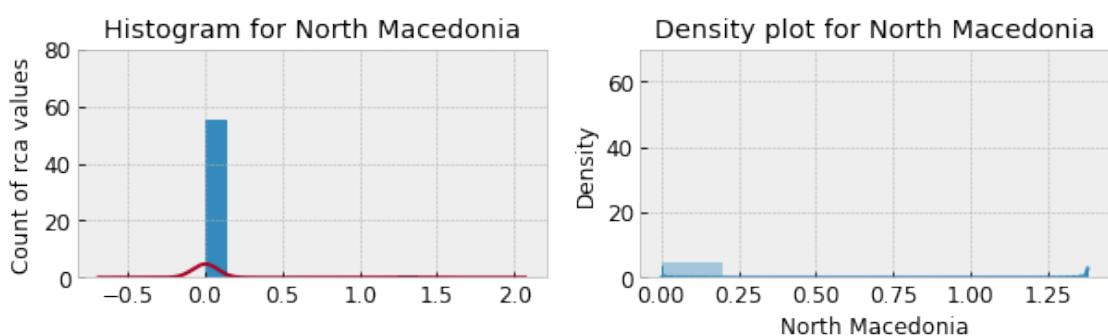
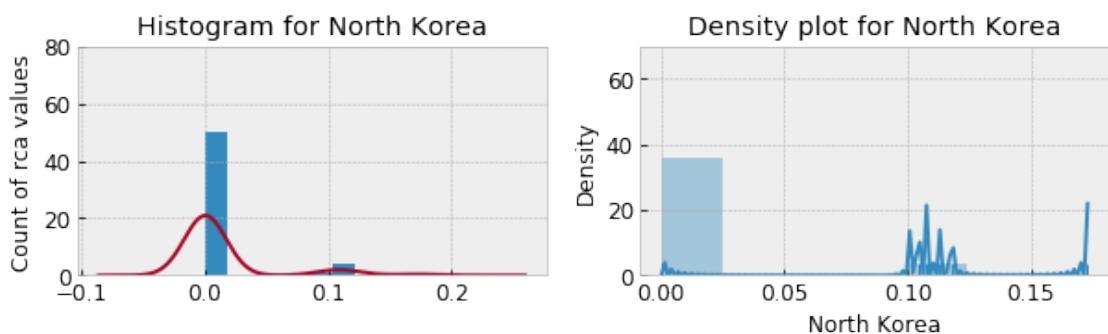
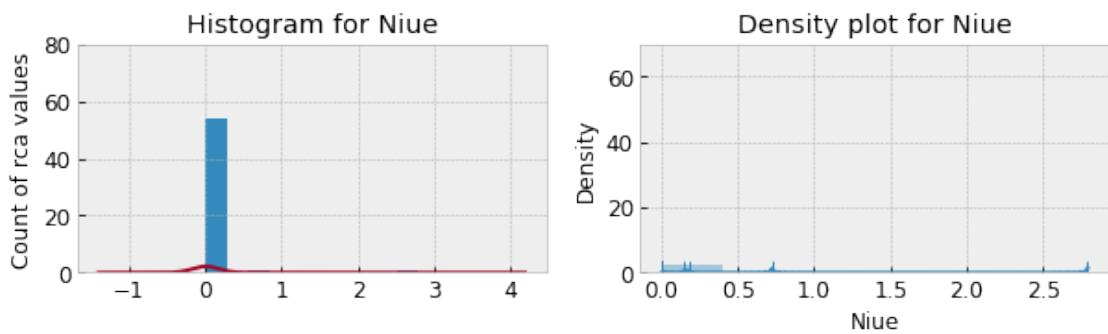
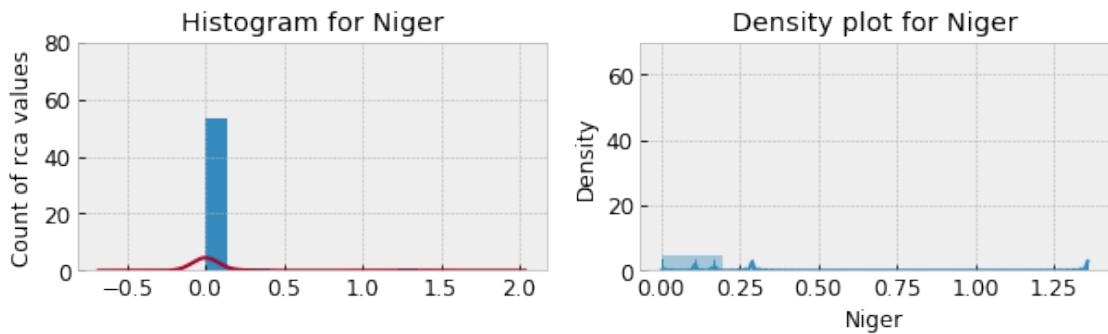


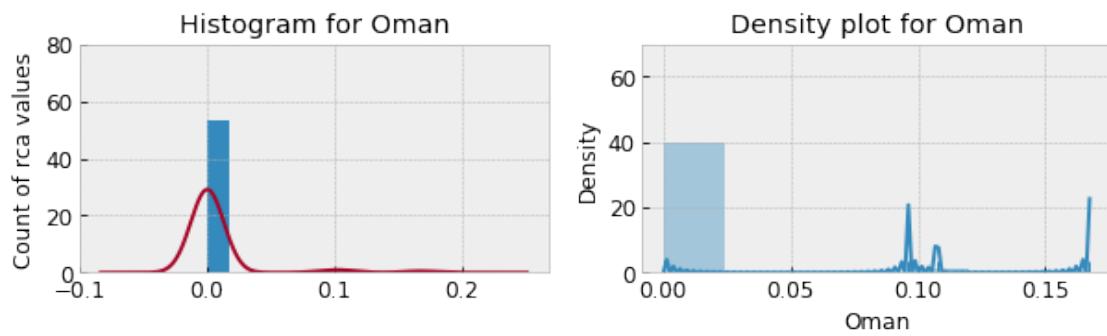
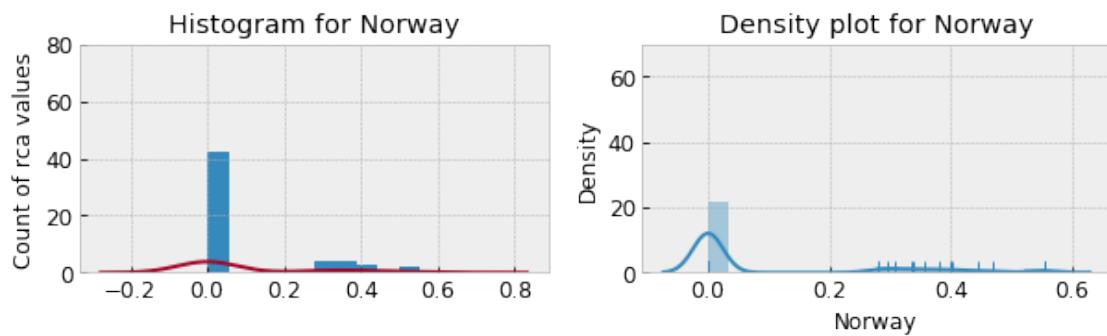
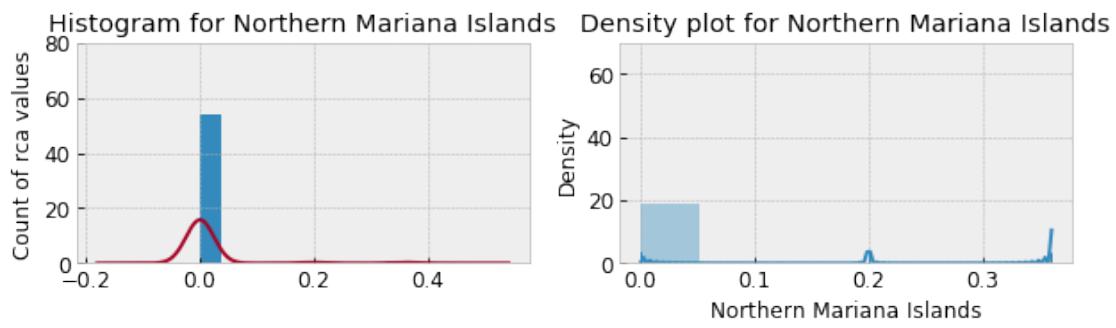


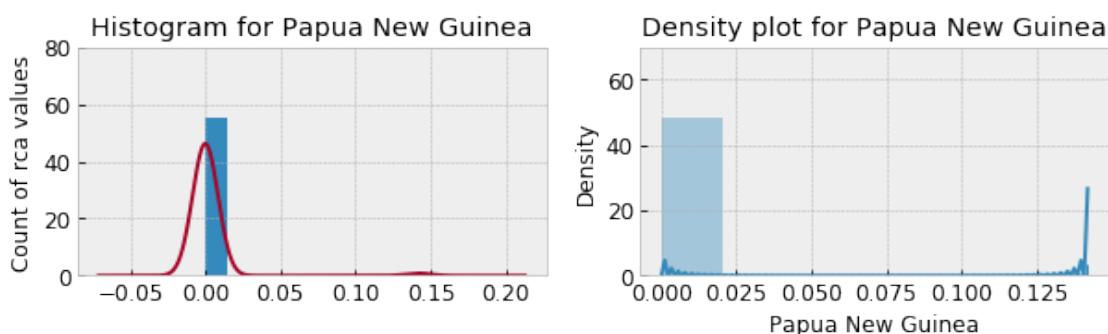
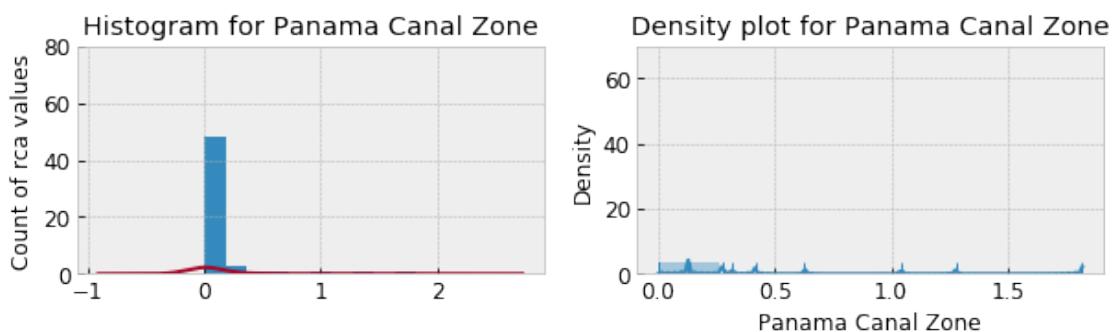
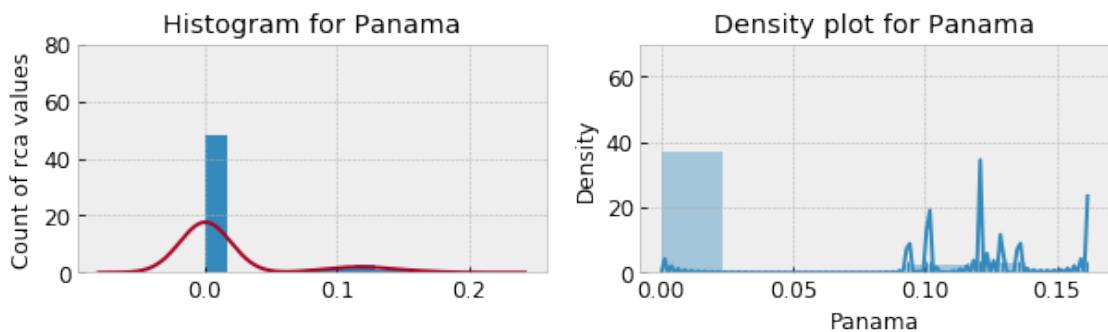
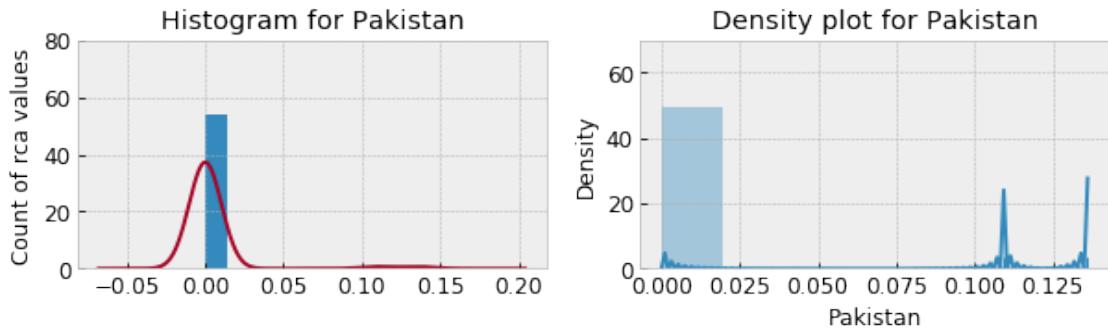


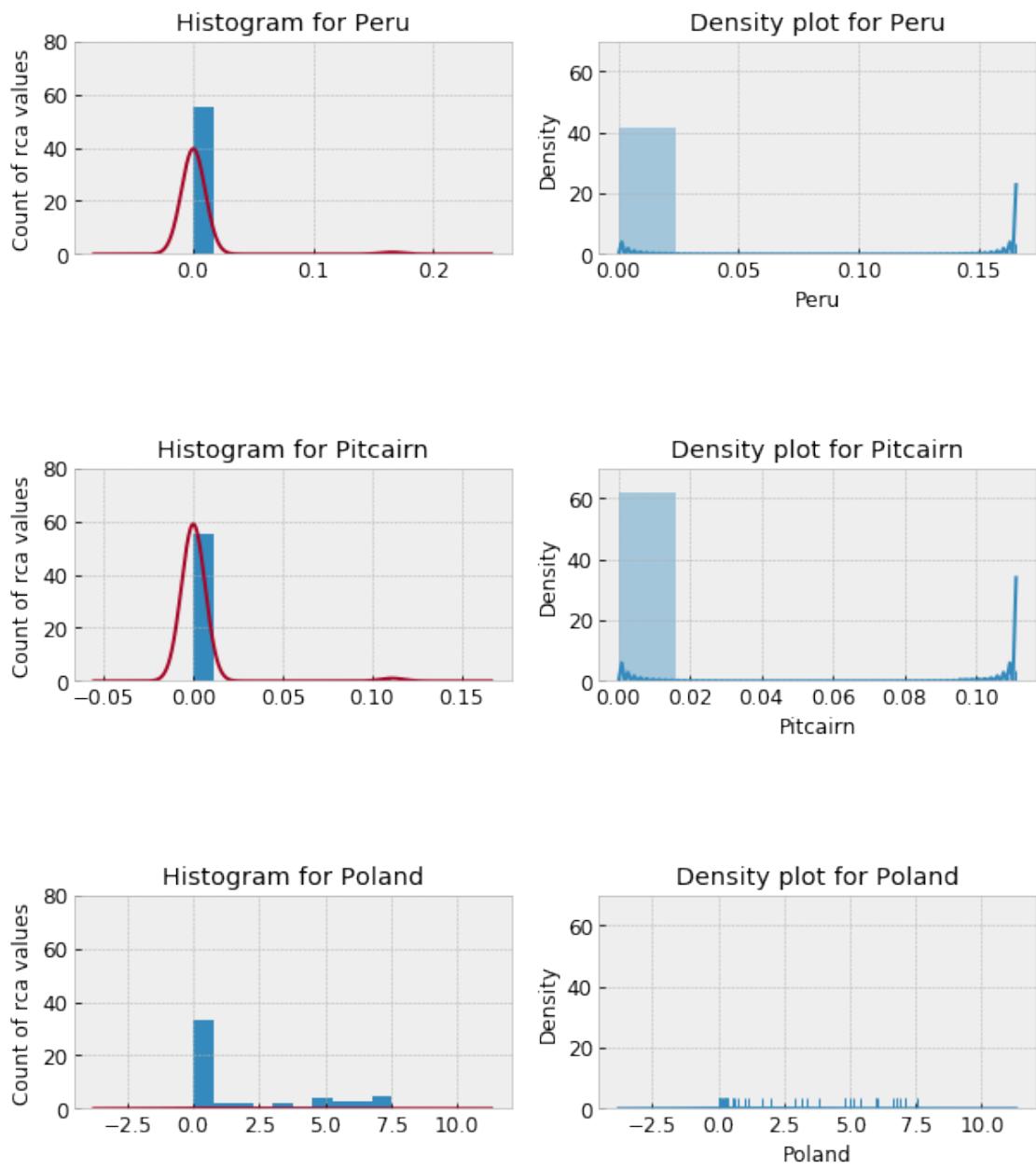


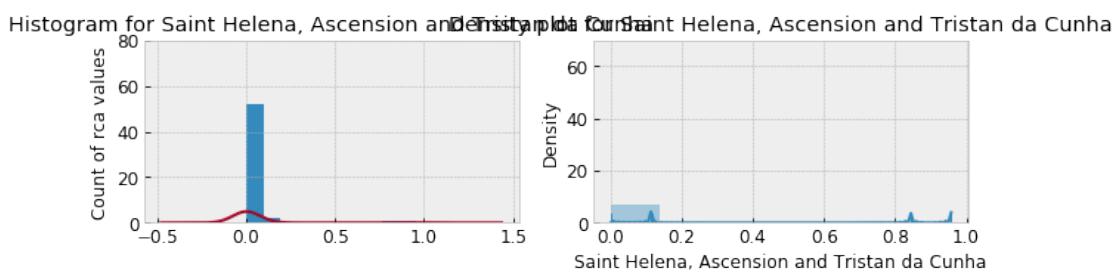
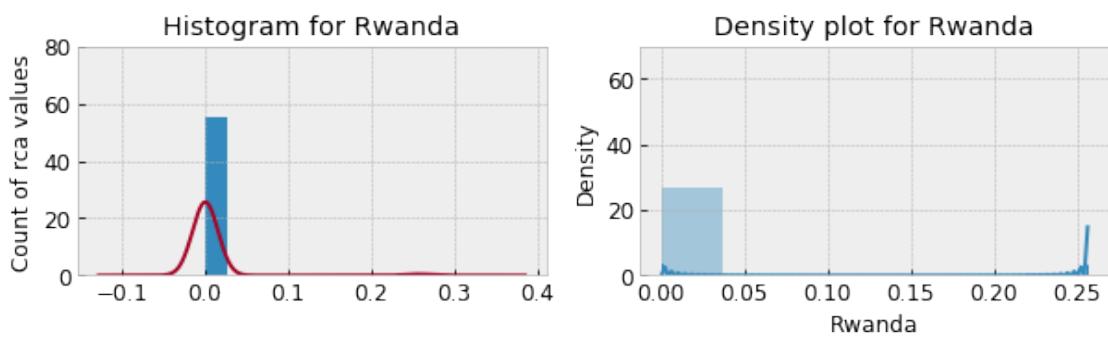
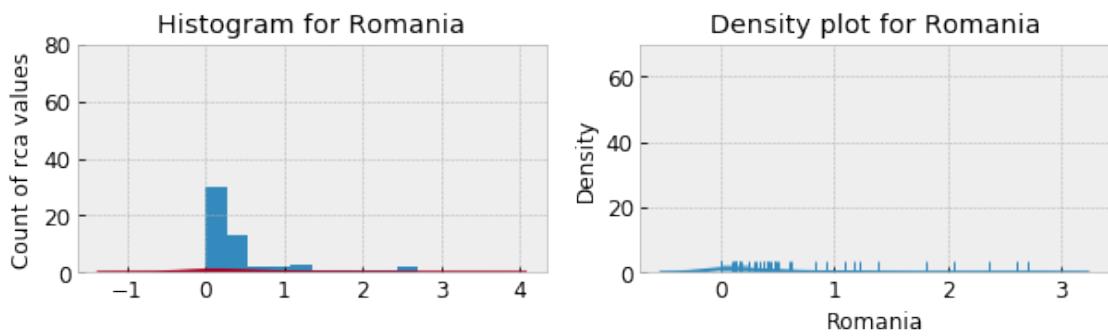
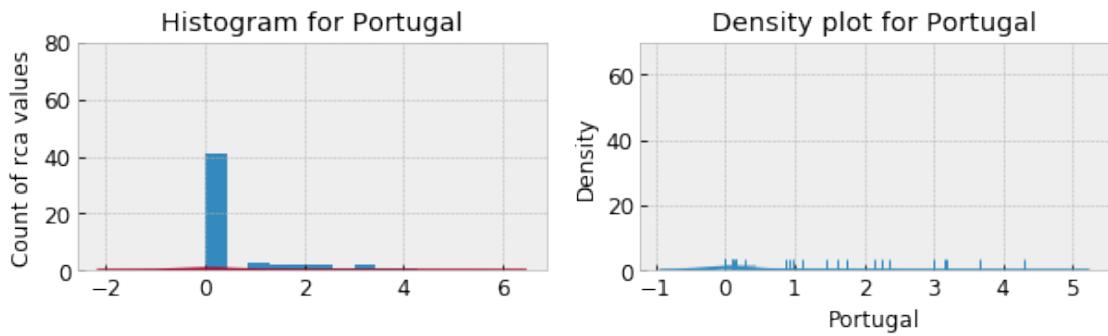


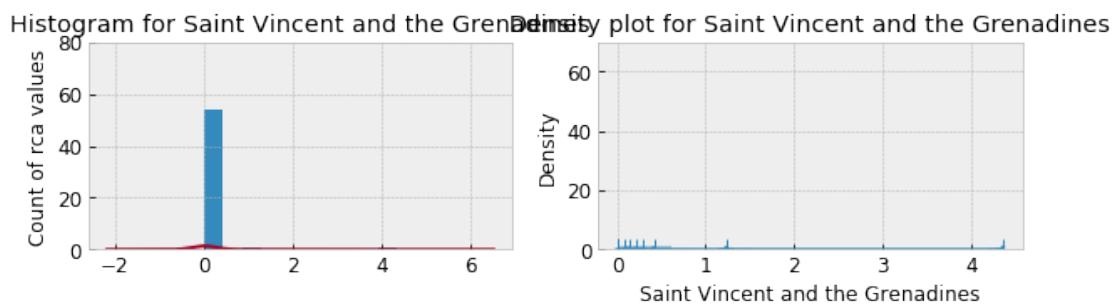
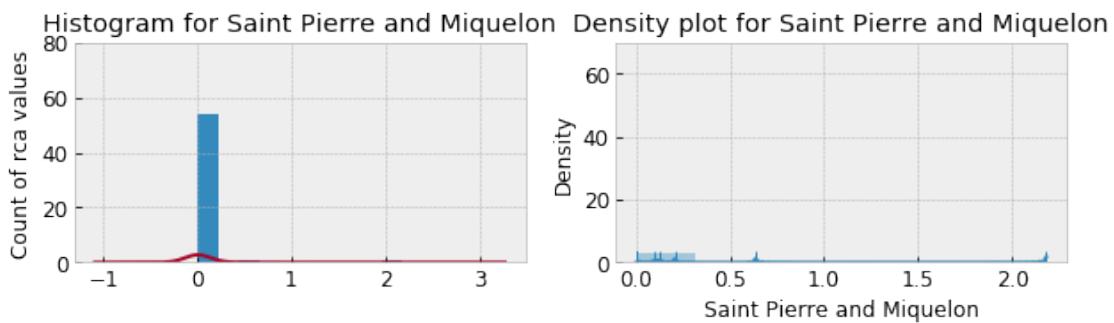
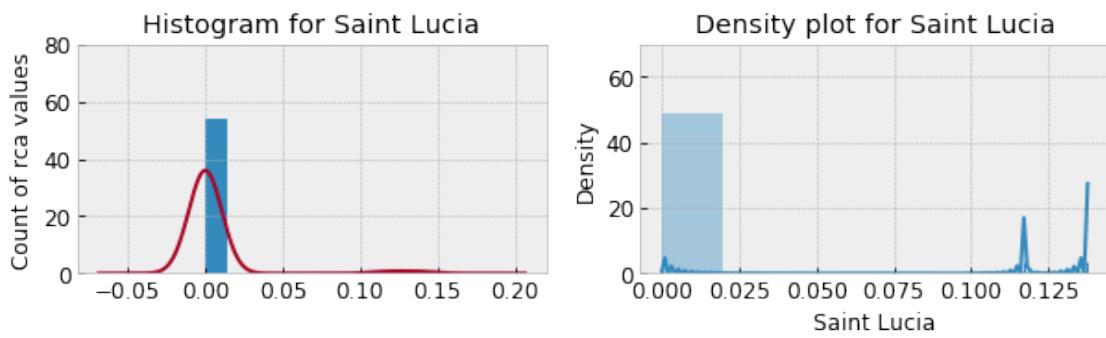
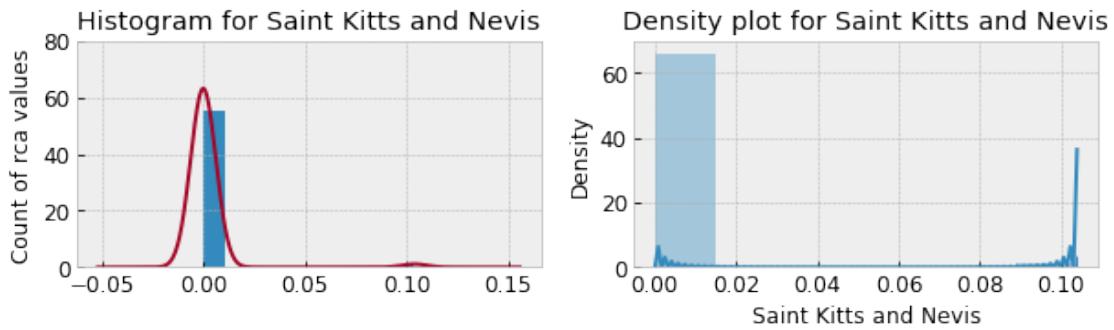


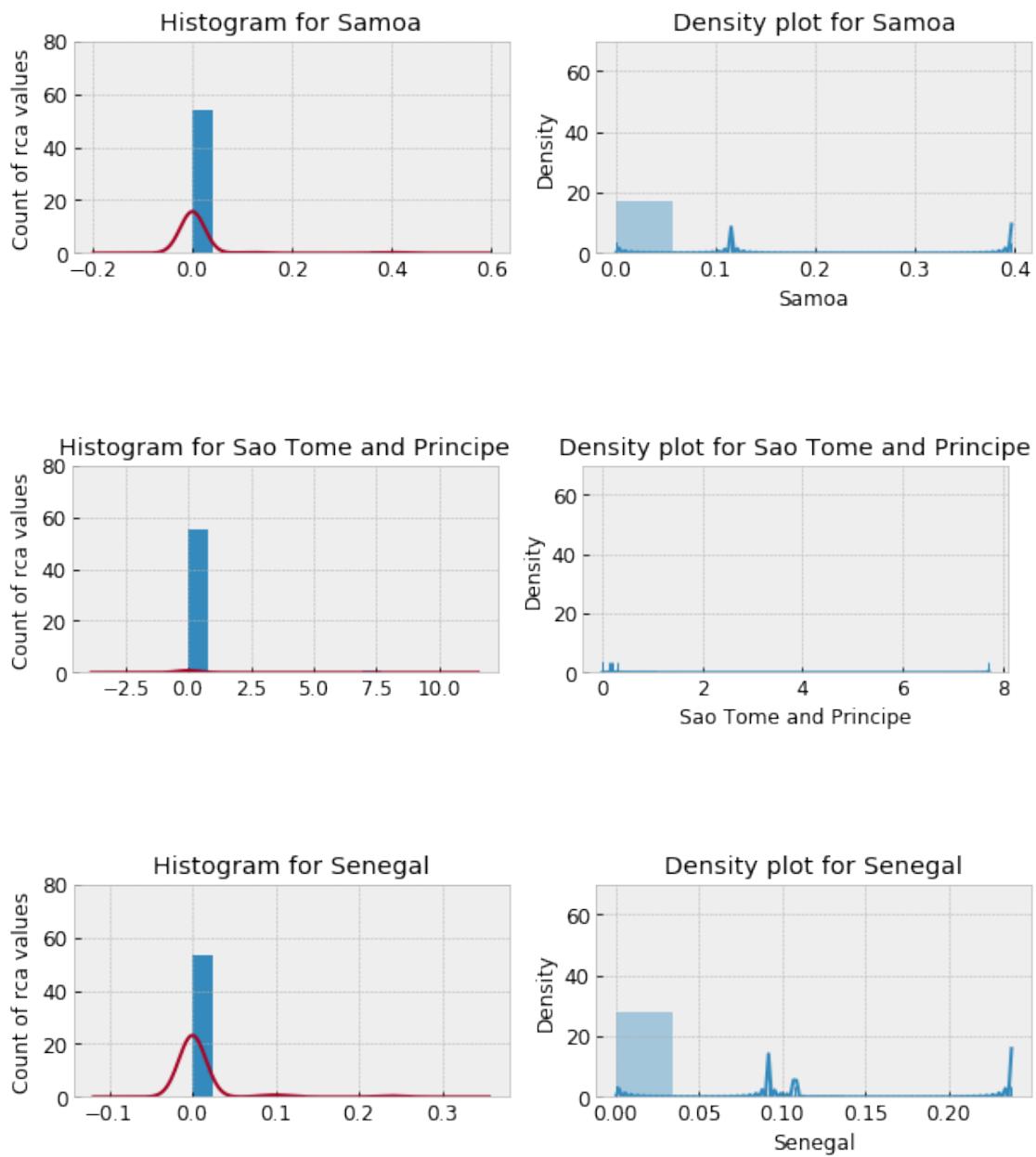


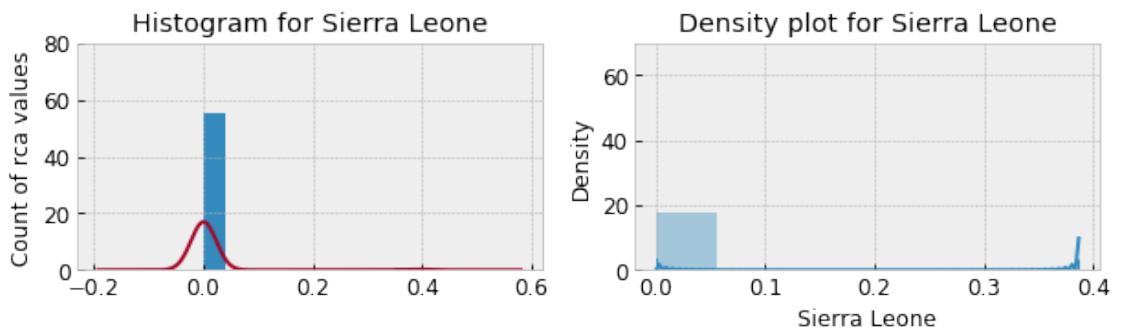
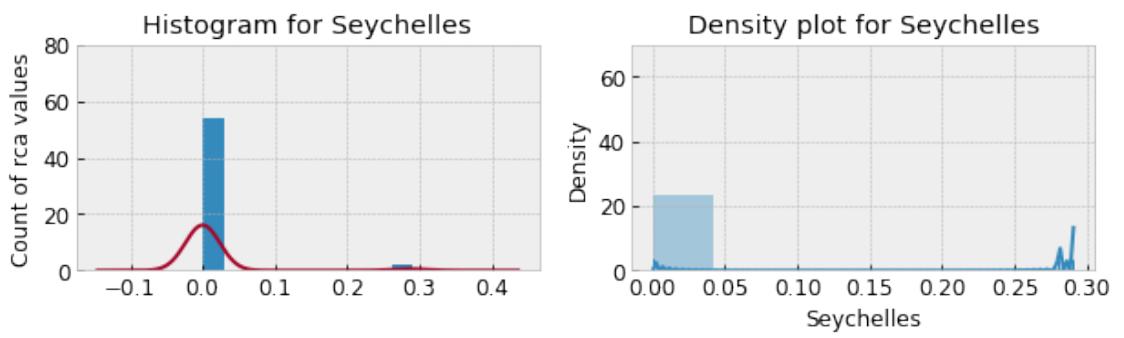
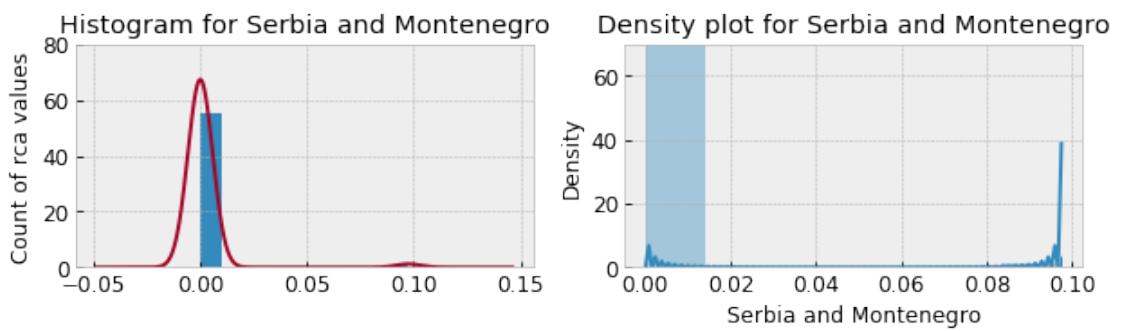
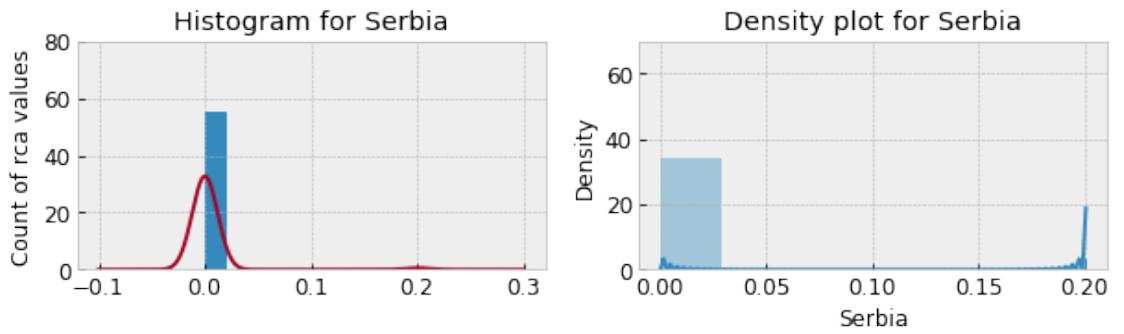


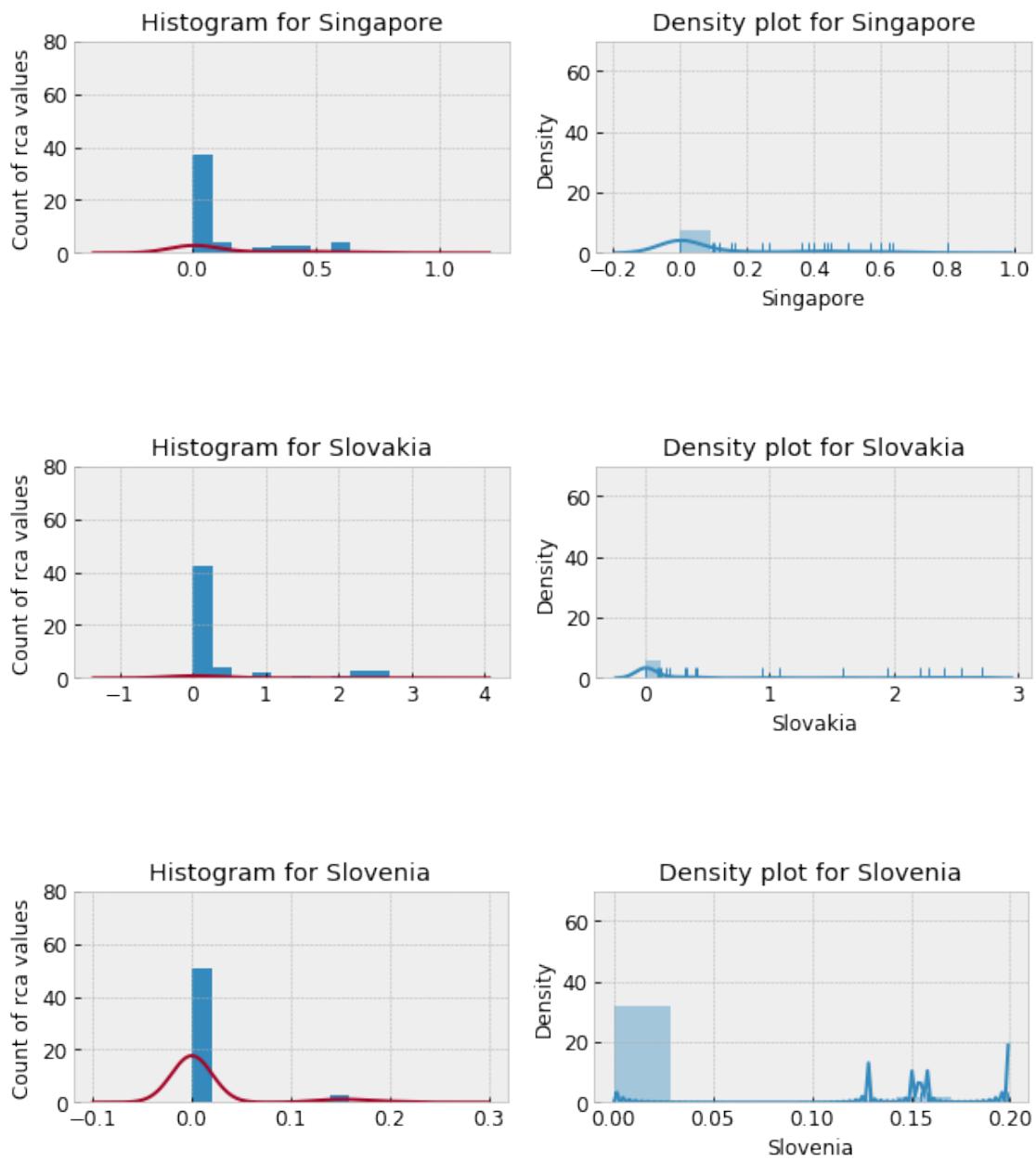


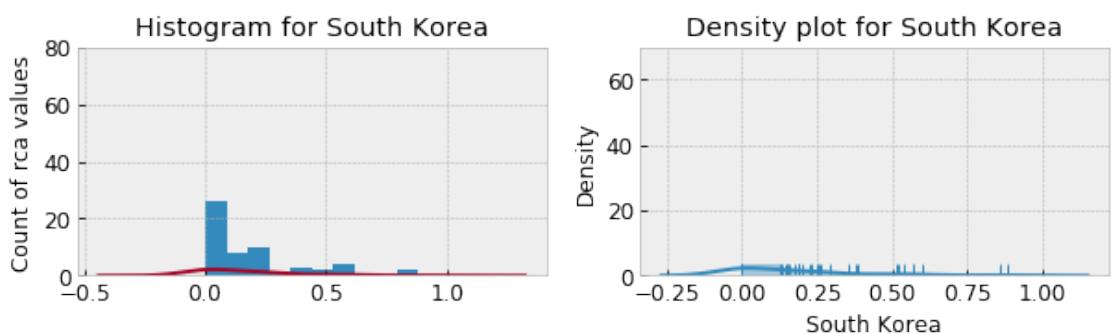
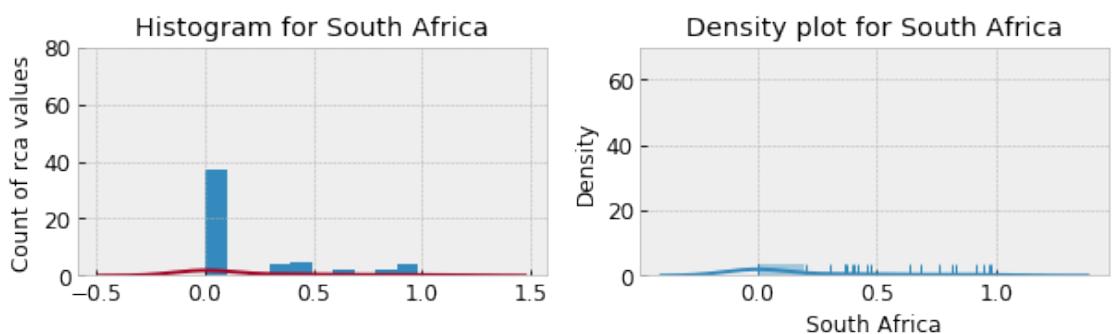
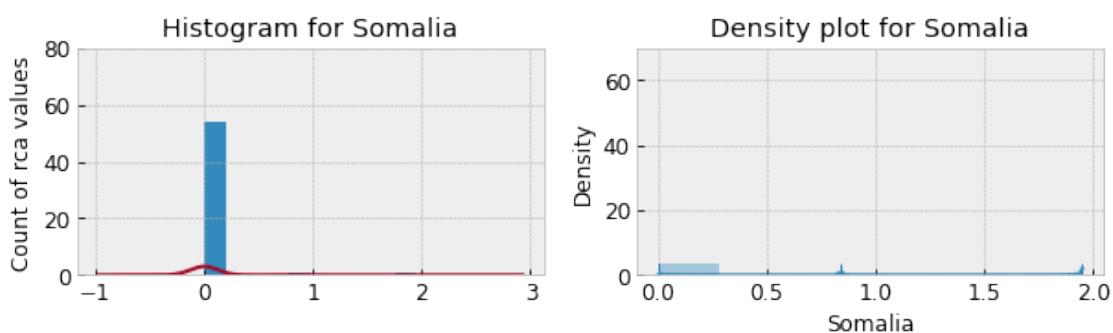
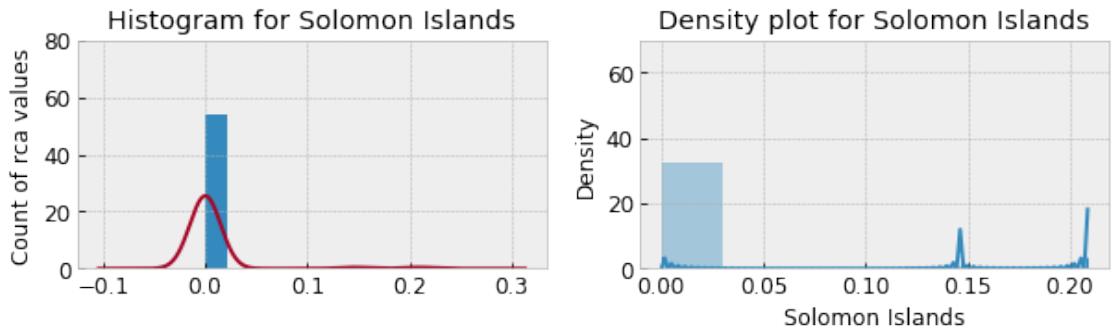


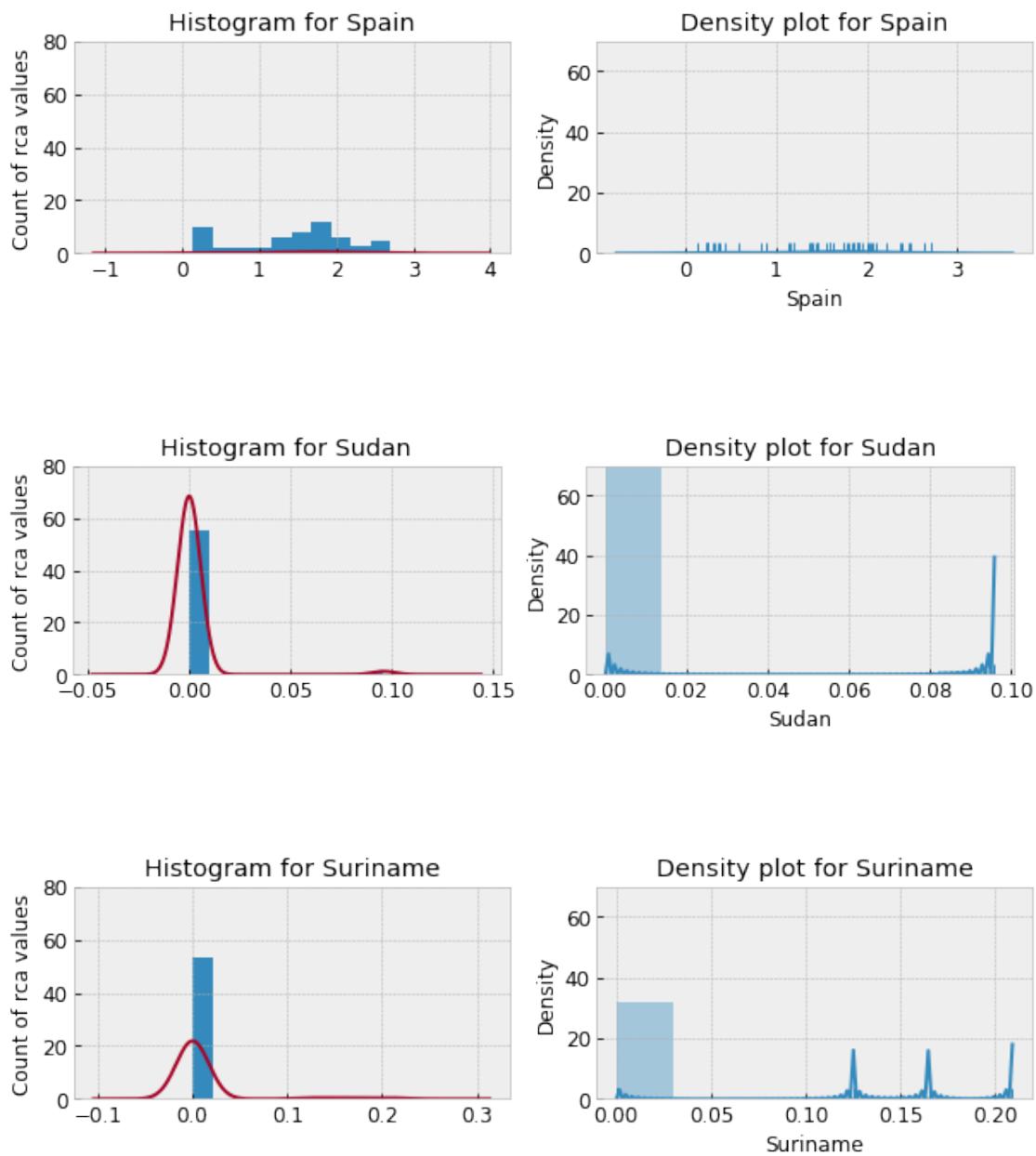


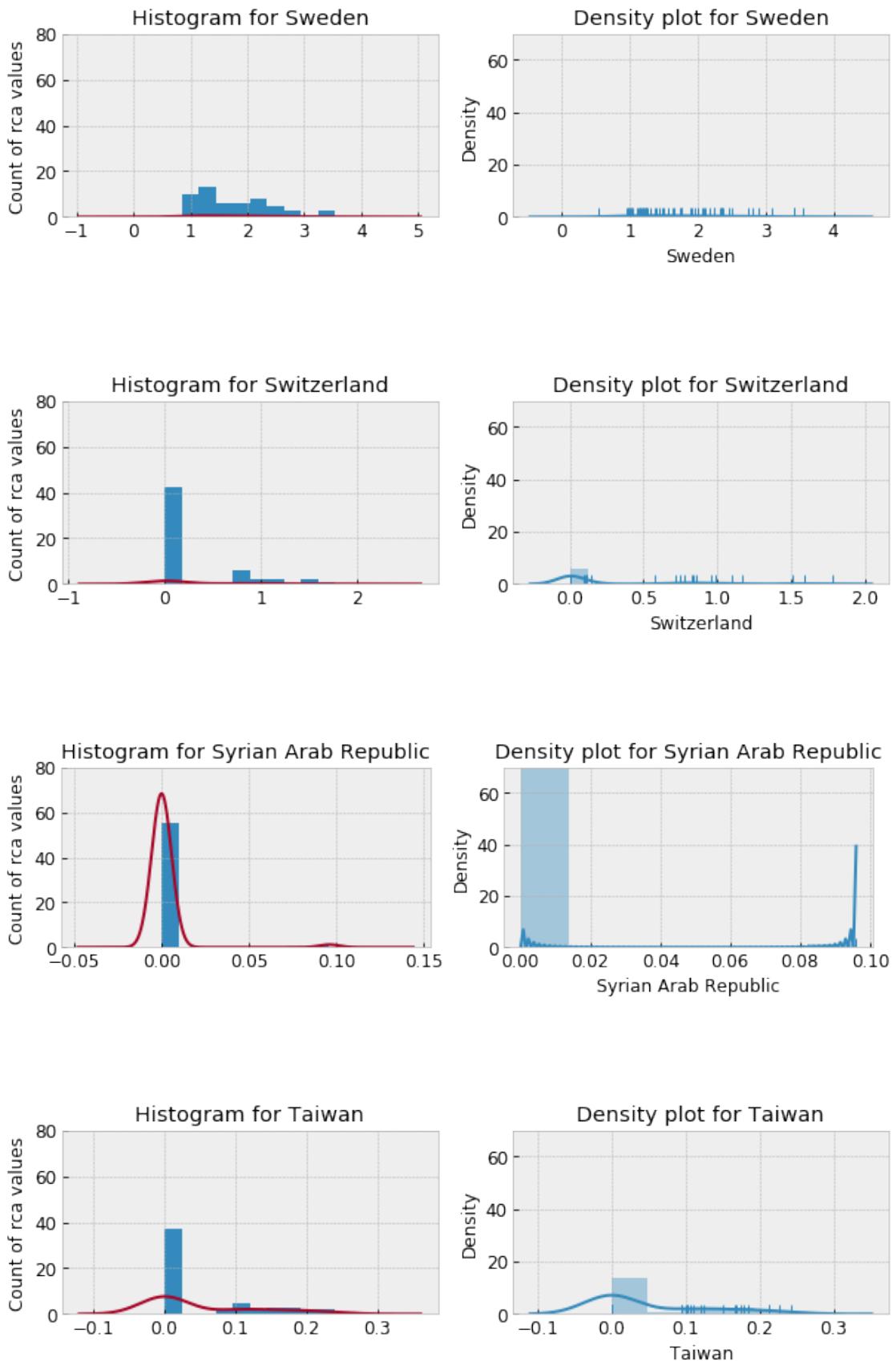


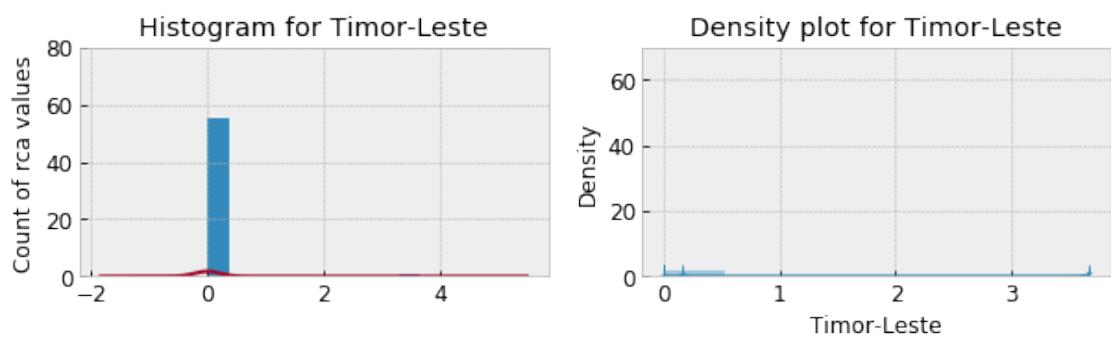
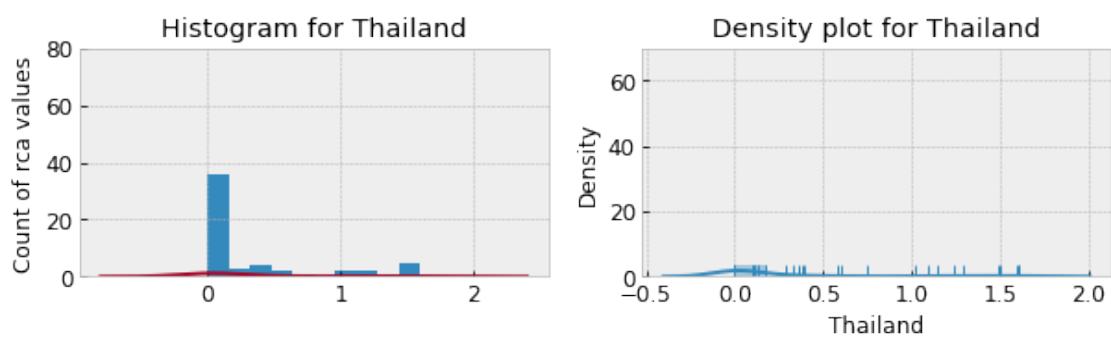
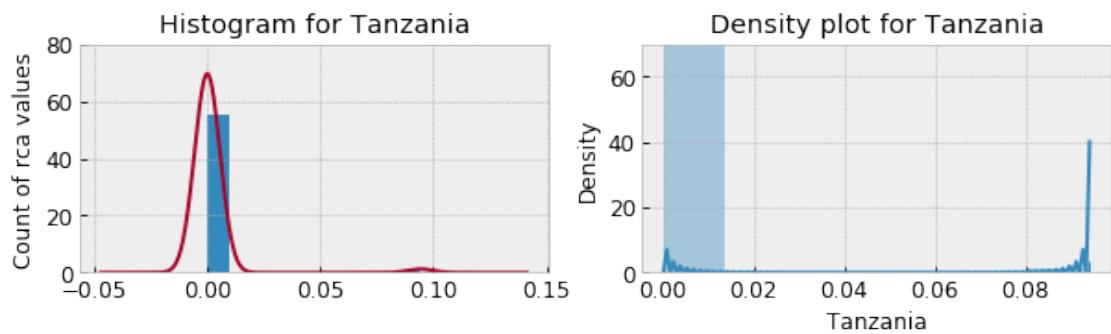


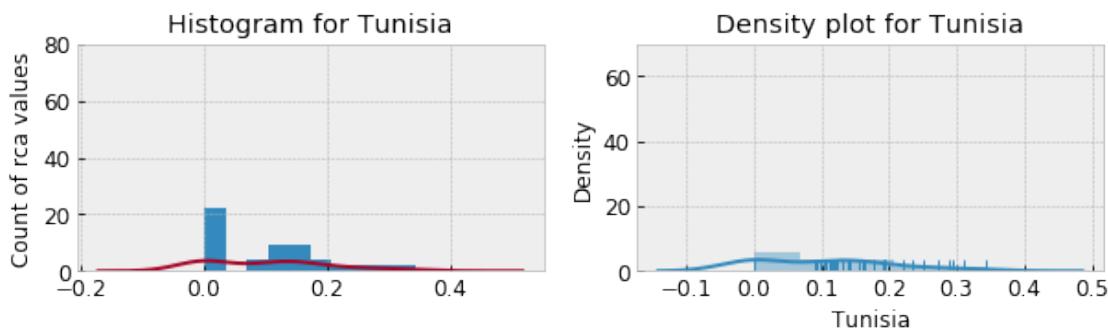
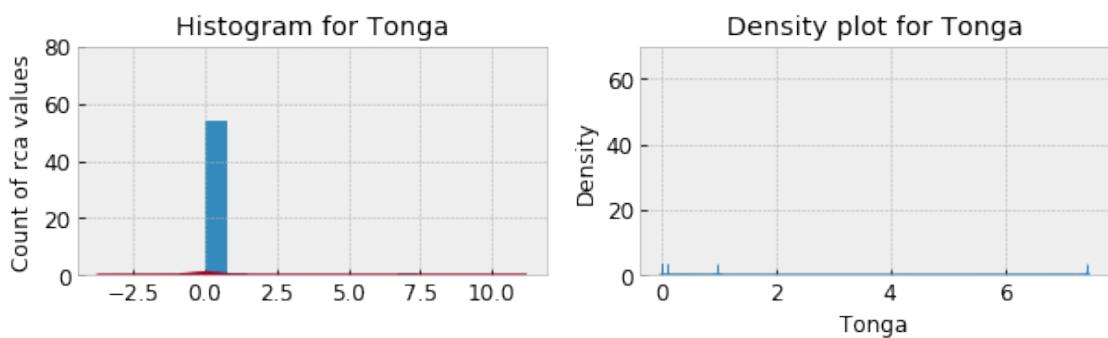
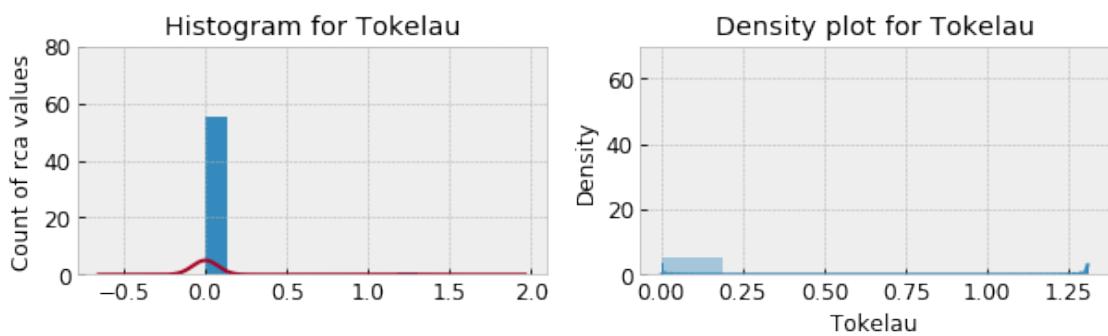
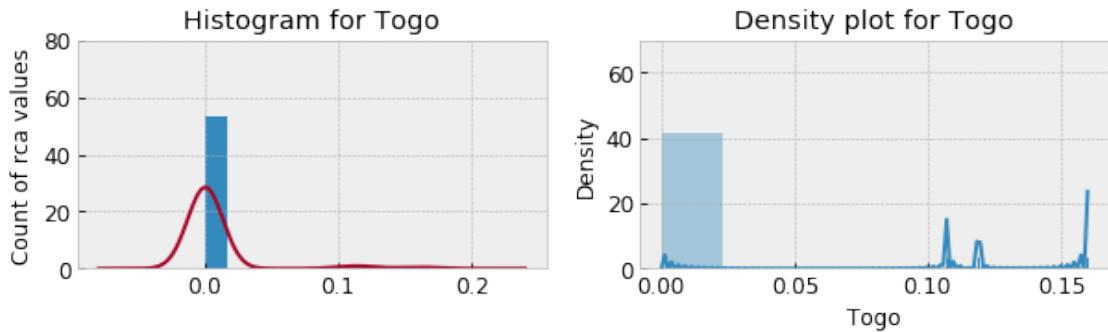


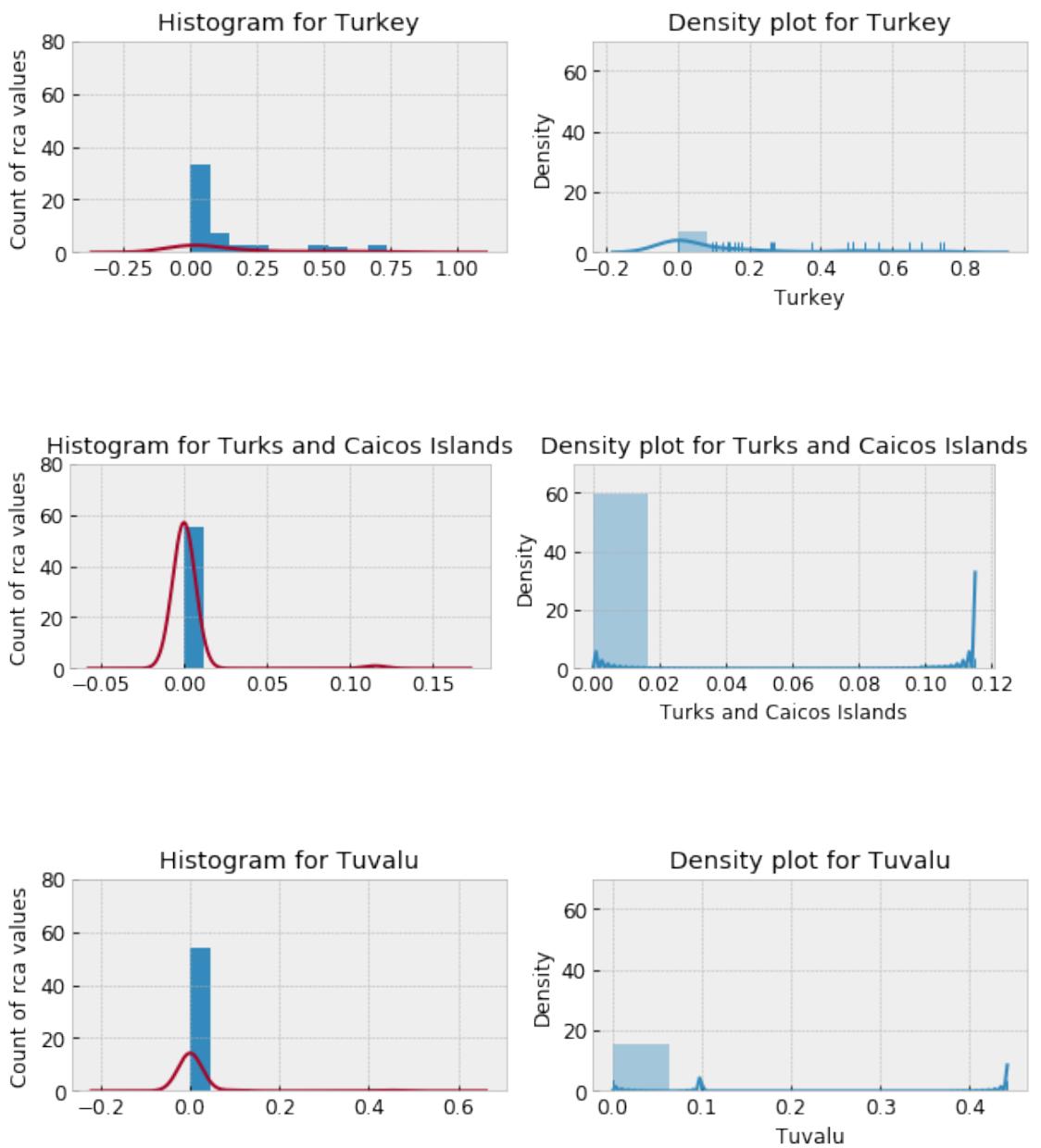


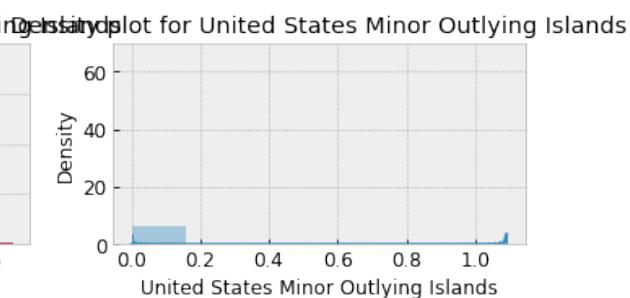
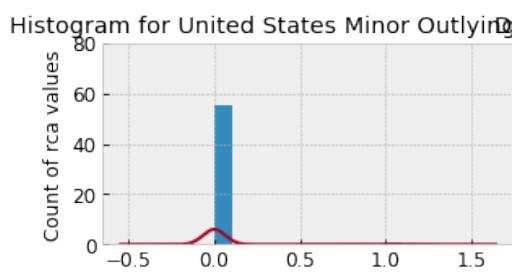
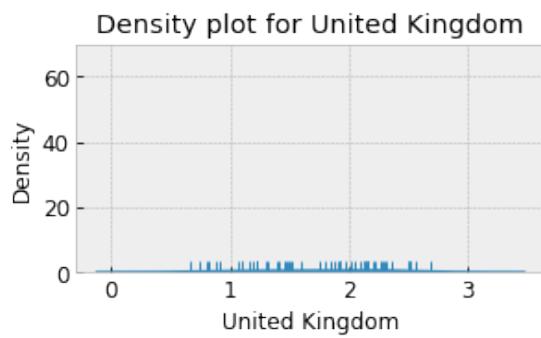
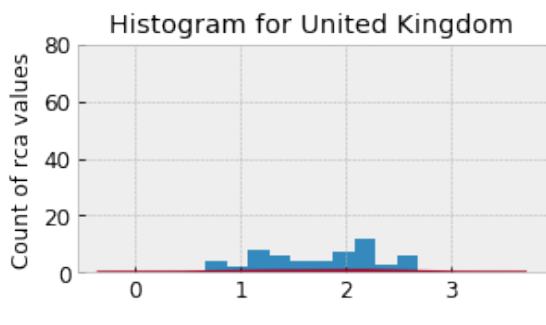
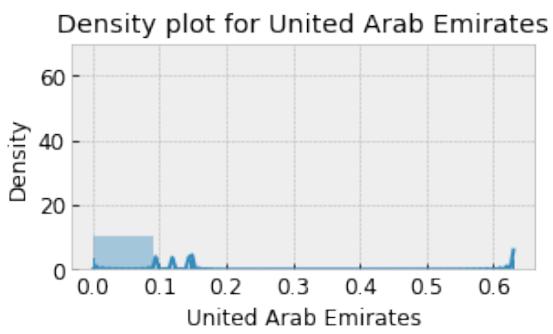
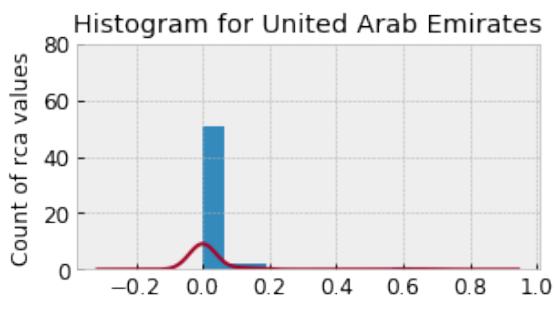
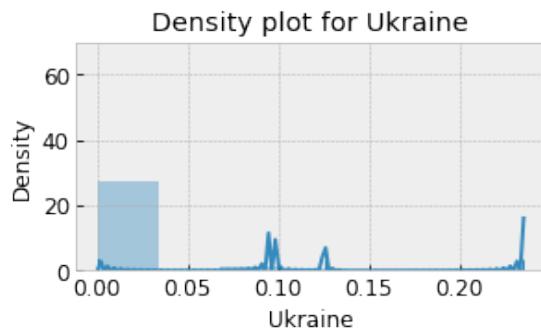
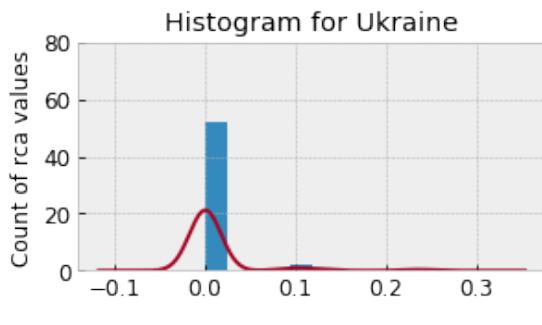


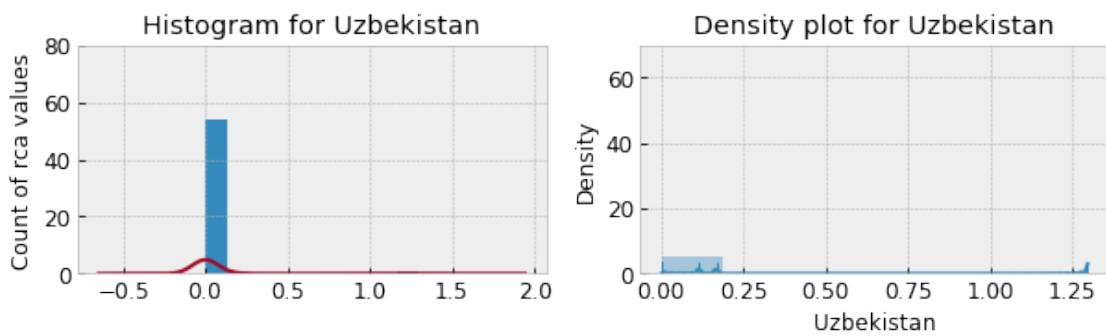
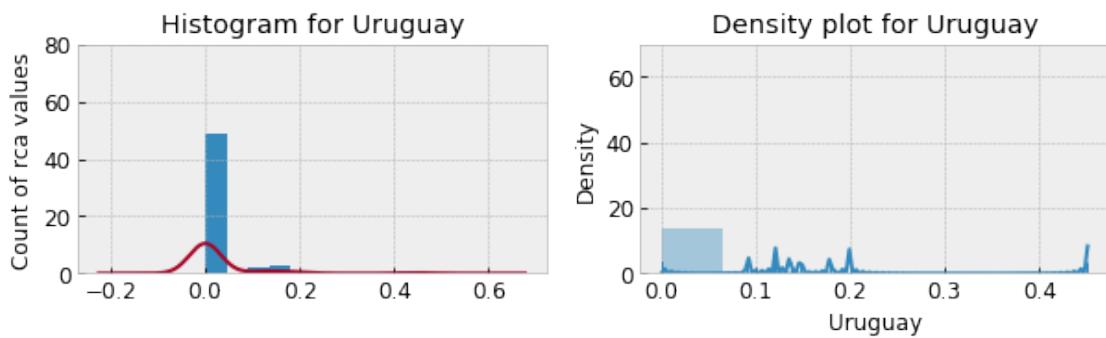
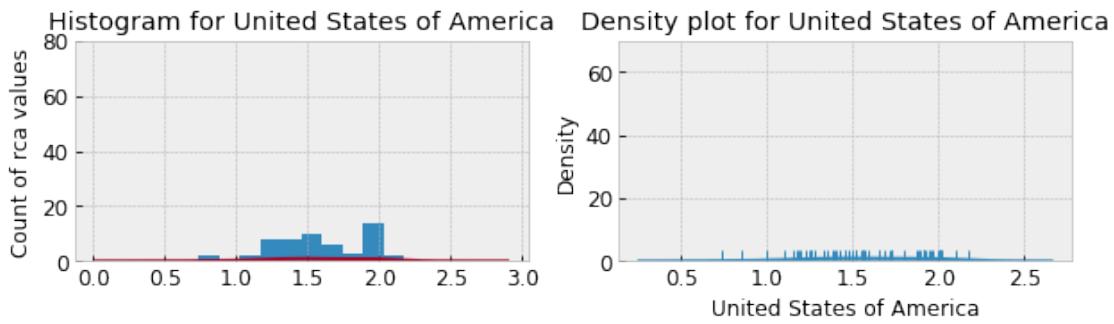


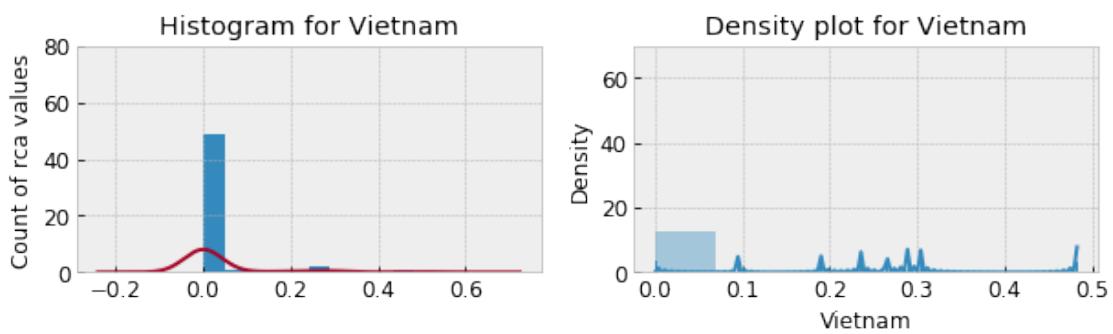
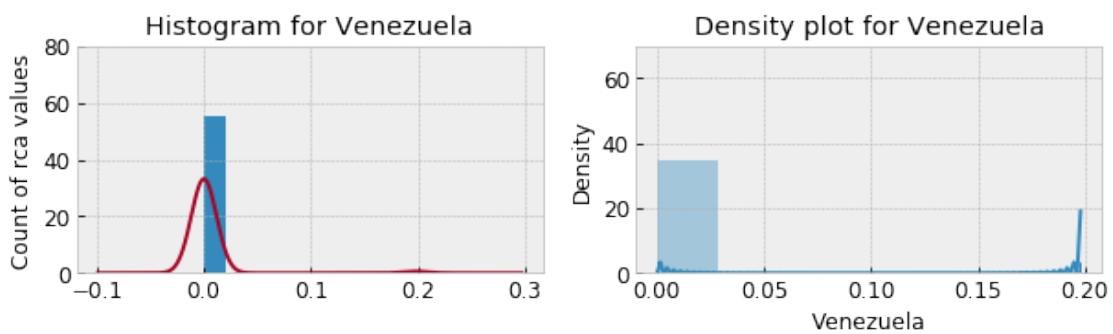
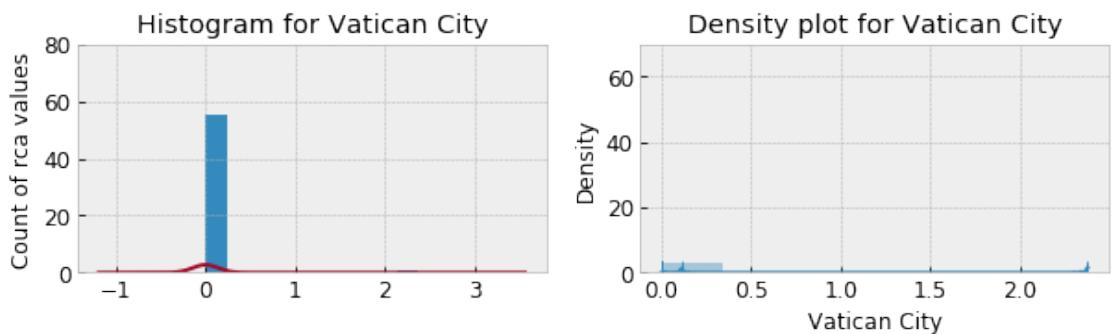
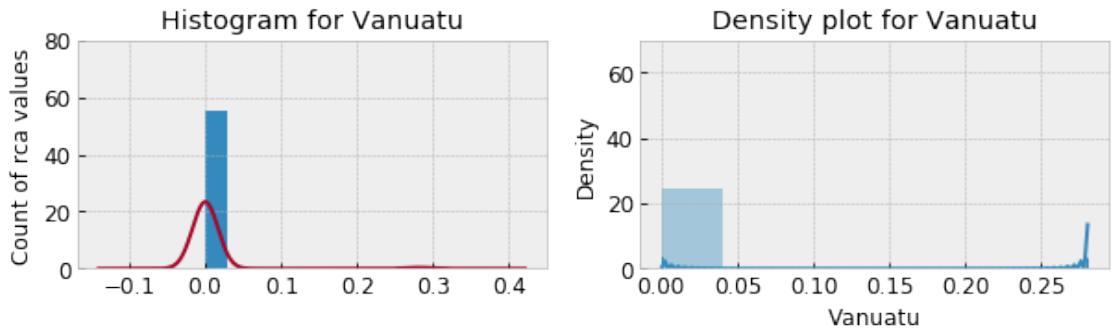


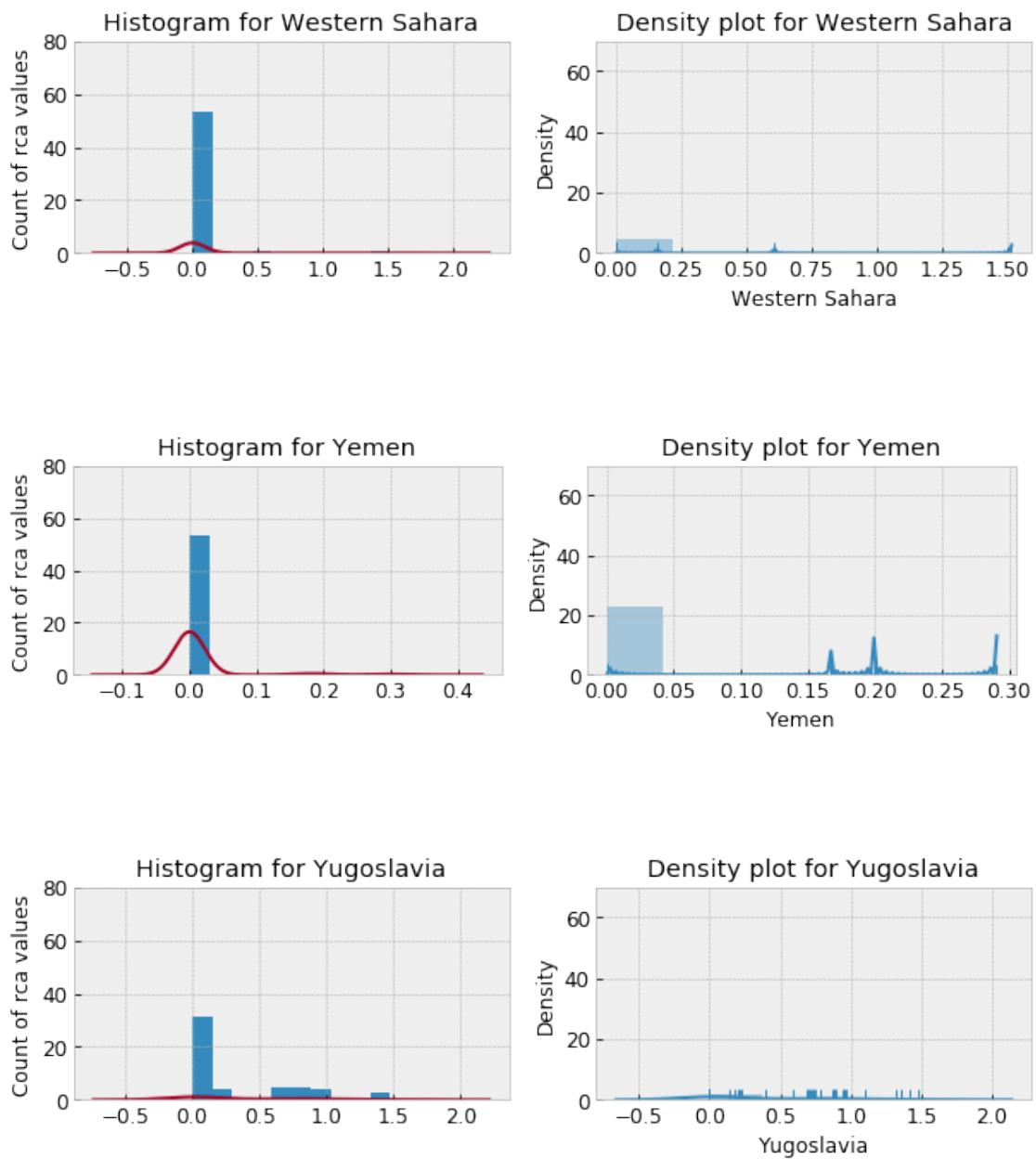


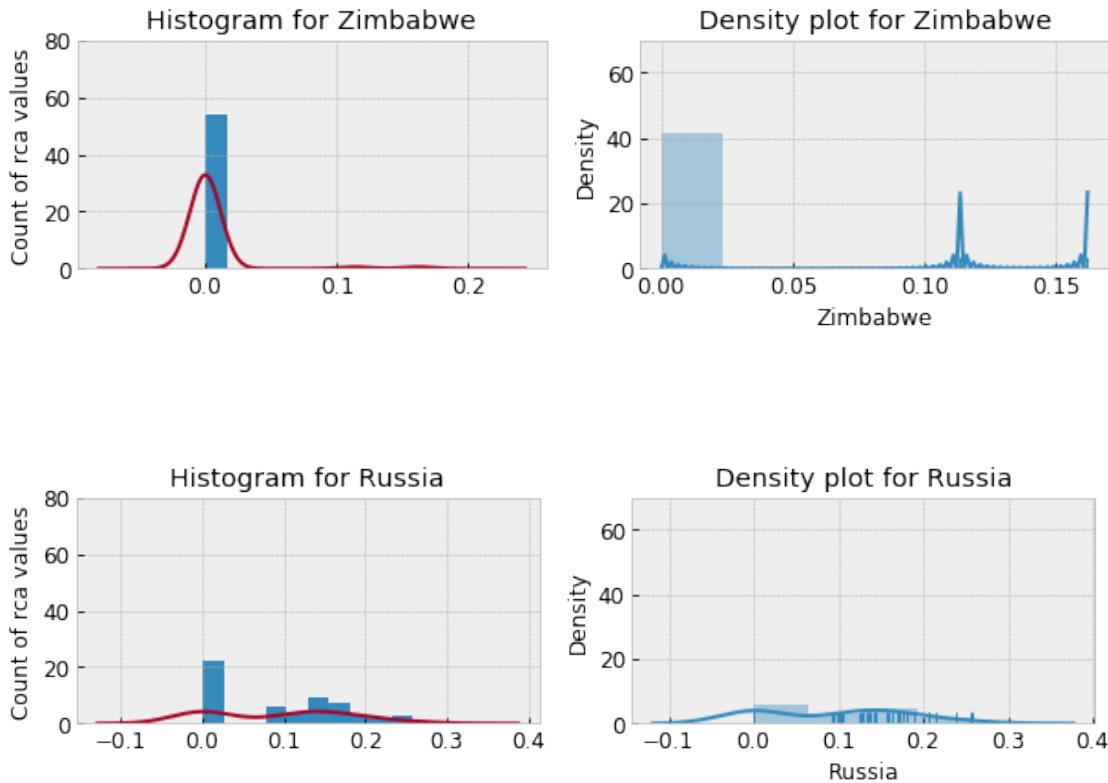












Box and Whisker Plots:

No high resolution in dataset to check: - what time series have similar median values across different years - whether there is a steady increase in the spread, or middle 50% of the data (boxes) over time - best model if considering seasonality

```
[174]: df_grps.info()
```

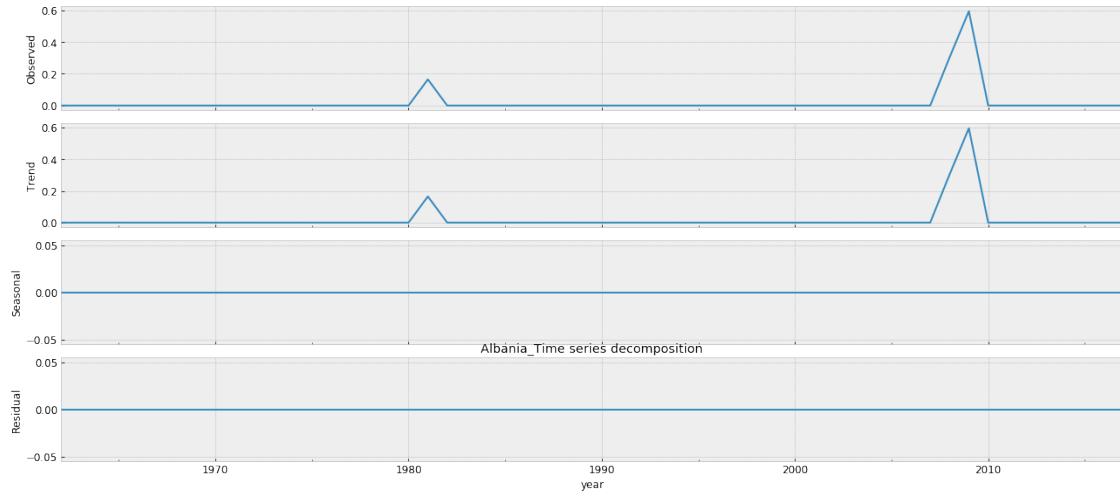
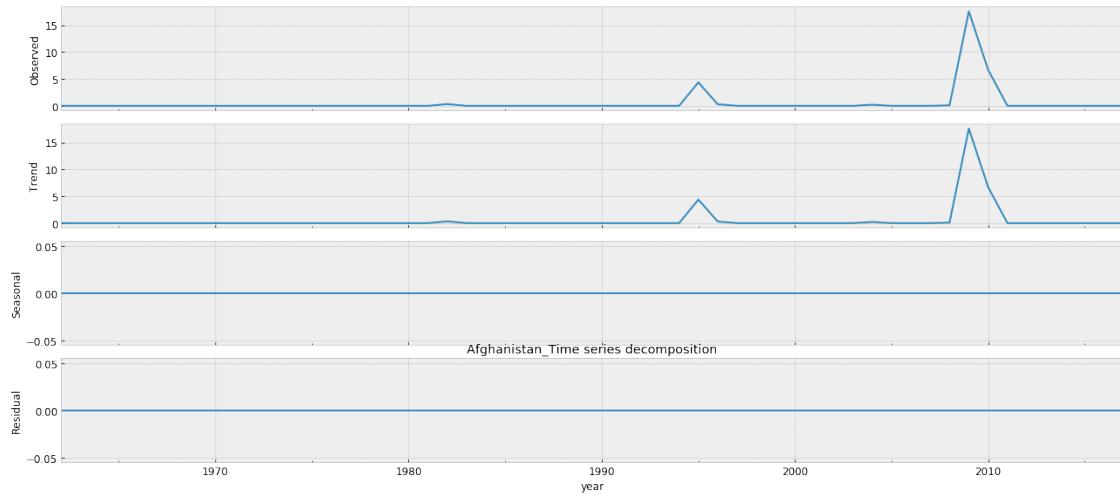
```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 56 entries, 1962-01-01 to 2017-01-01
Columns: 183 entries, Afghanistan to Russia
dtypes: float64(183)
memory usage: 80.5 KB
```

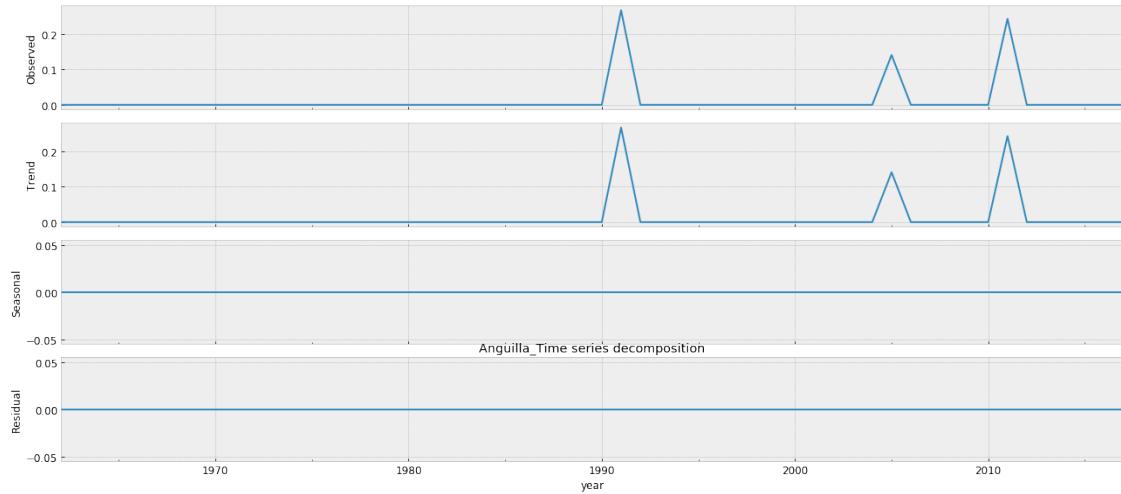
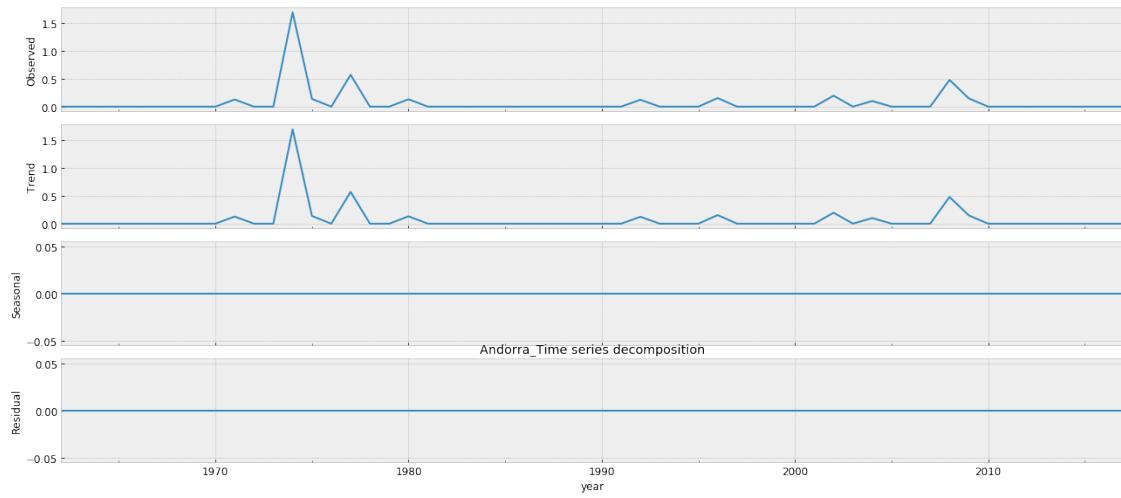
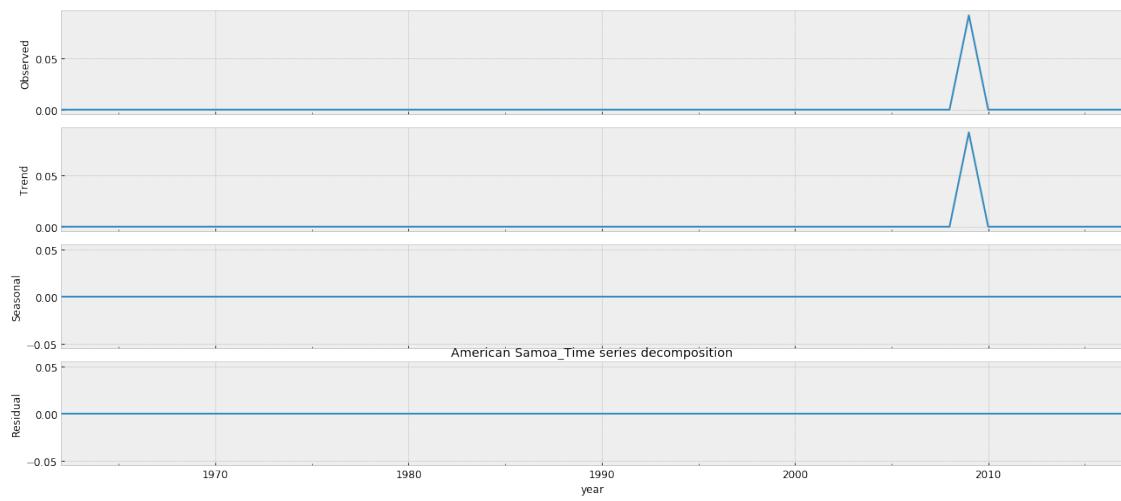
Decomposing using statsmodel: - statsmodels was used to perform a decomposition of the time series. - Decomposition deconstructs a time series into several components, each representing one of the underlying categories of patterns. - Statsmodels reveals the trend, seasonality, and residual components of our data.

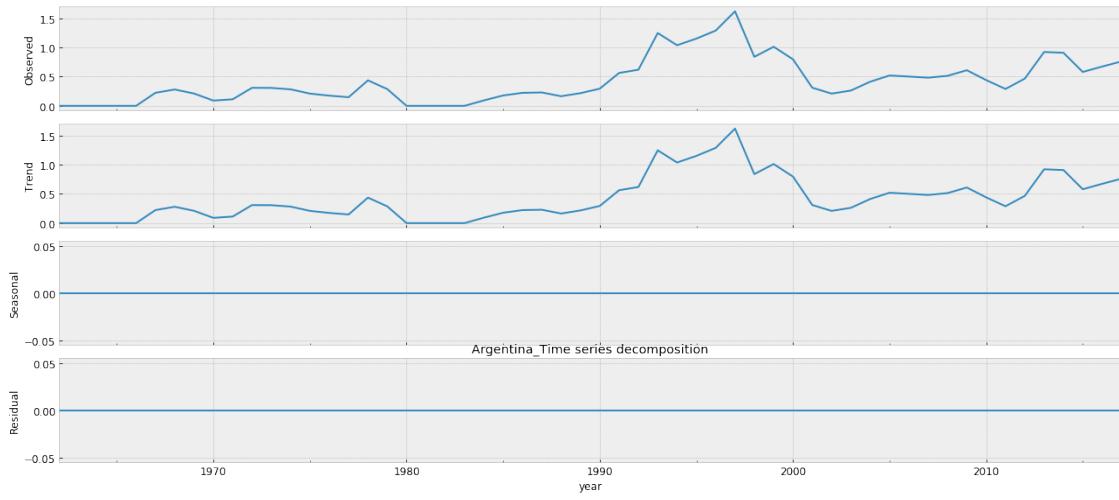
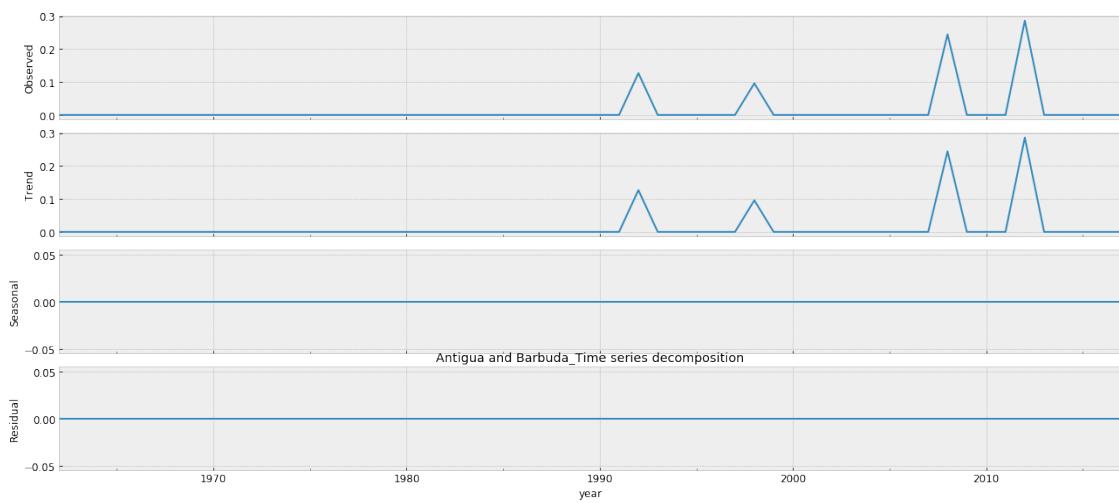
```
[175]: countries = df_grps.columns
```

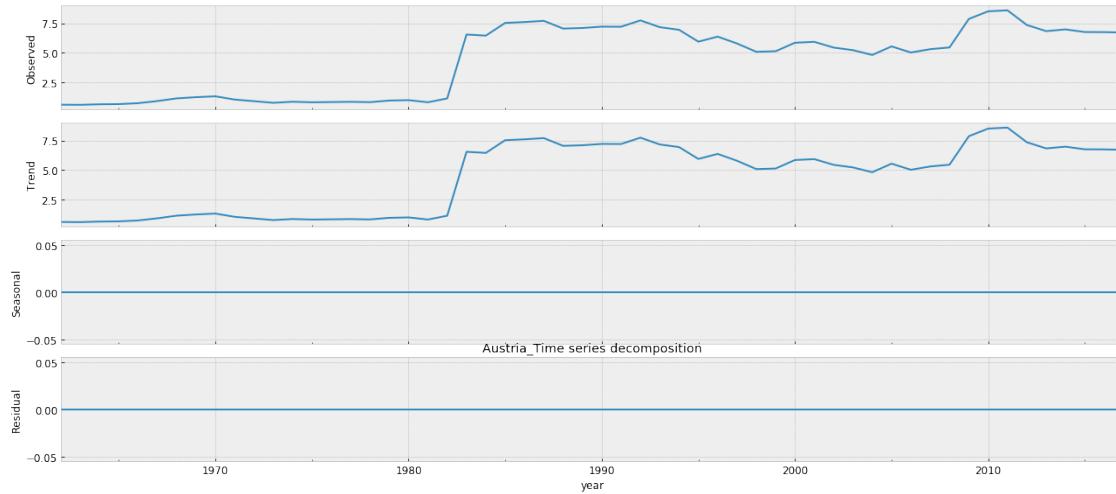
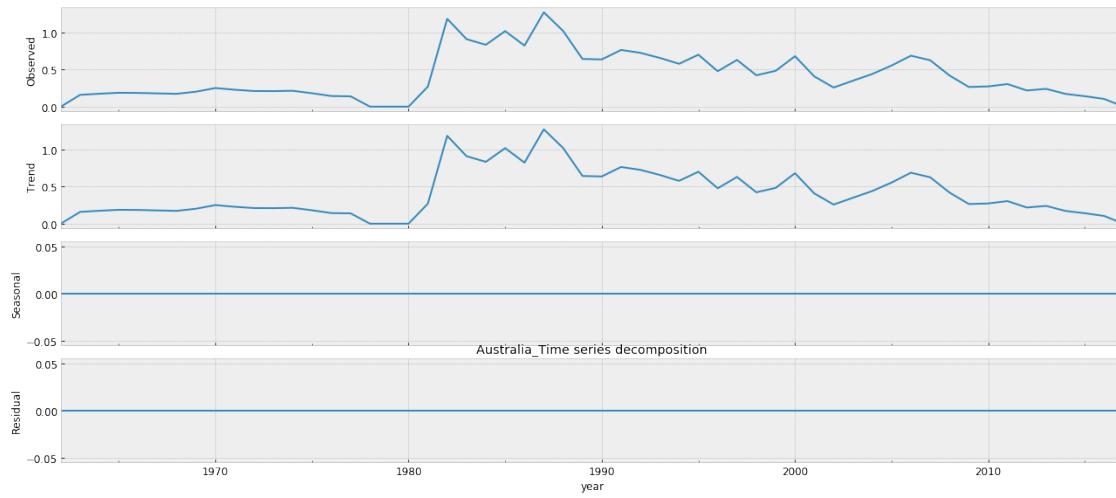
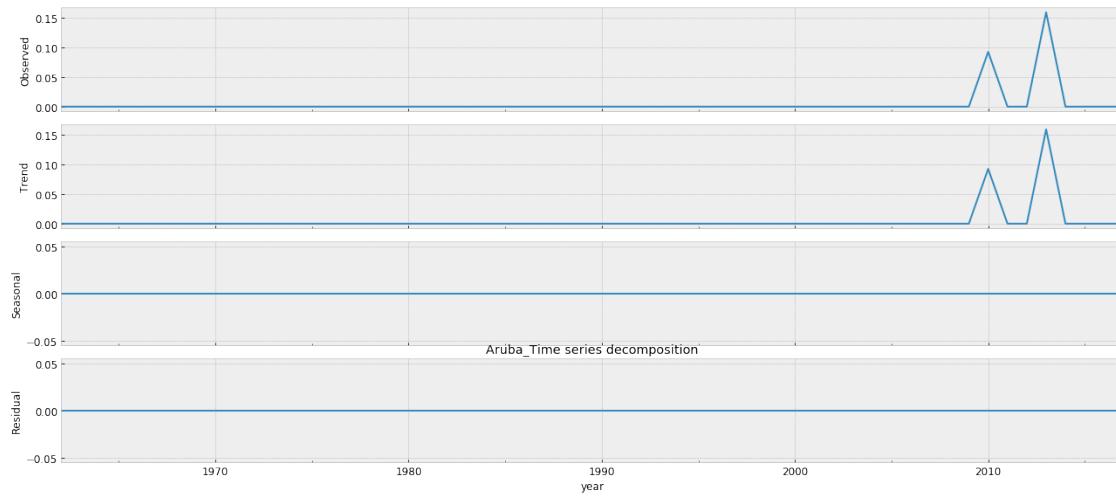
Decomposition

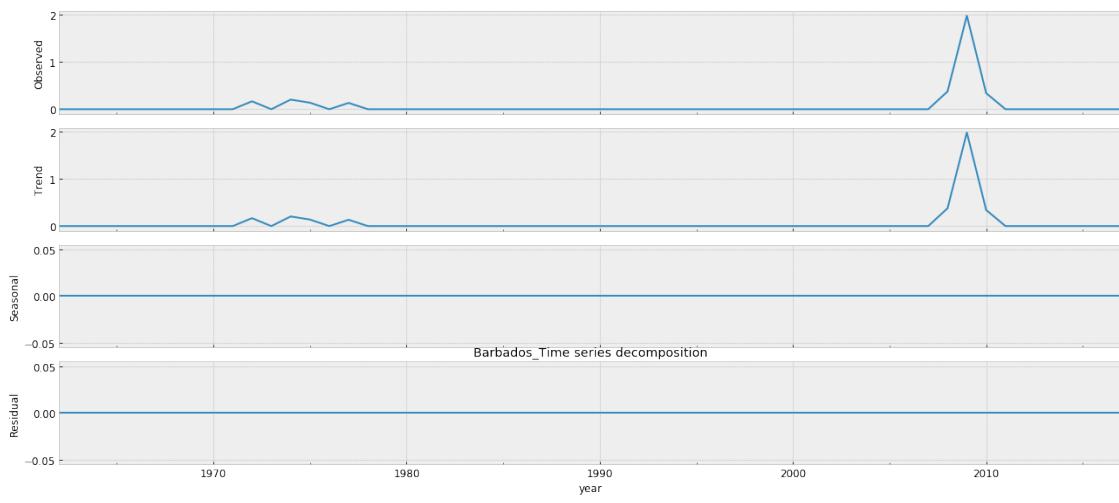
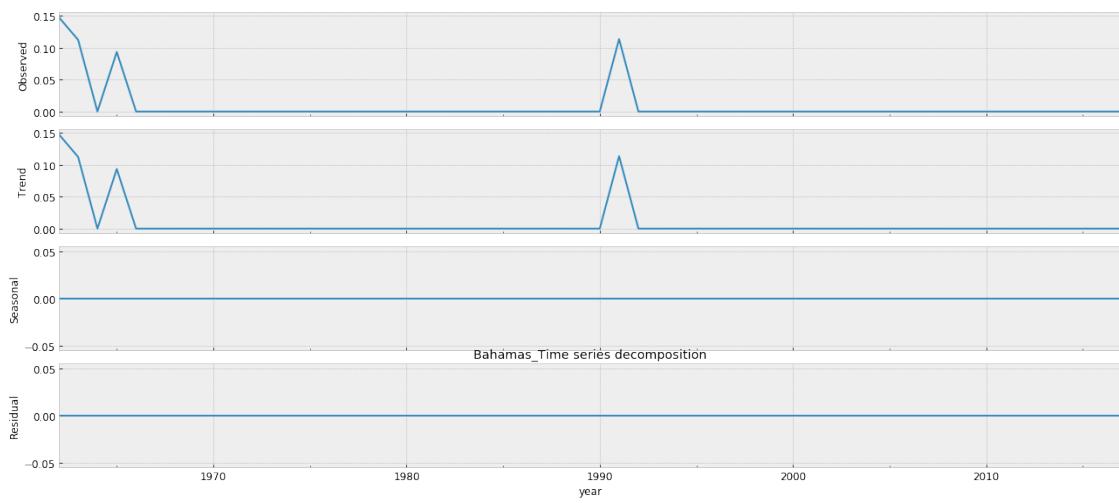
```
[176]: for country in countries:
    savefile = save_images(df_grps, 'decomposition')
    decomposition = sm.tsa.seasonal_decompose(df_grps[country], model='additive')
    fig = decomposition.plot()
    plt.title(f'{df_grps[country].name}_Time series decomposition', loc='center')
    rcParams['figure.figsize'] = (18, 8)
    plt.savefig(str(savefile))
    plt.show()
```

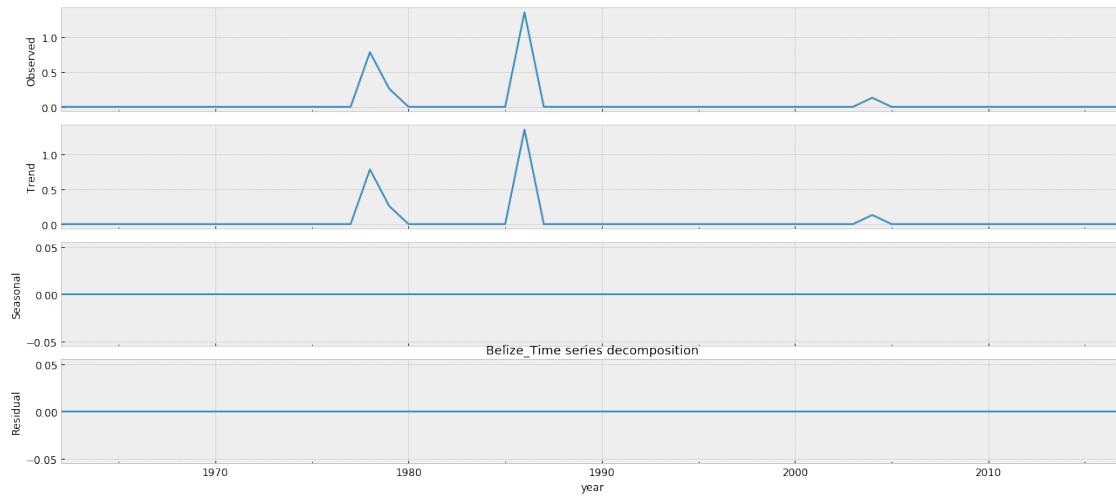
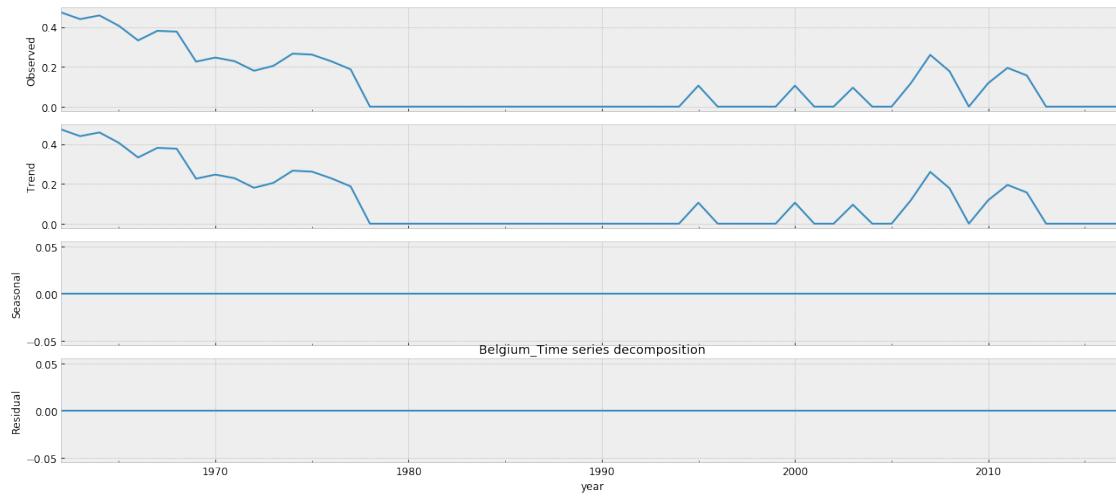
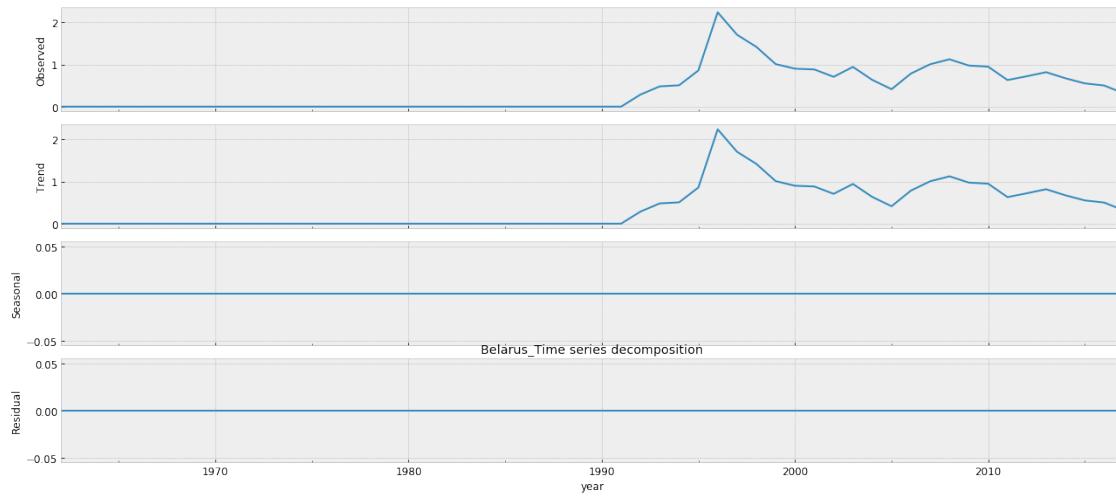


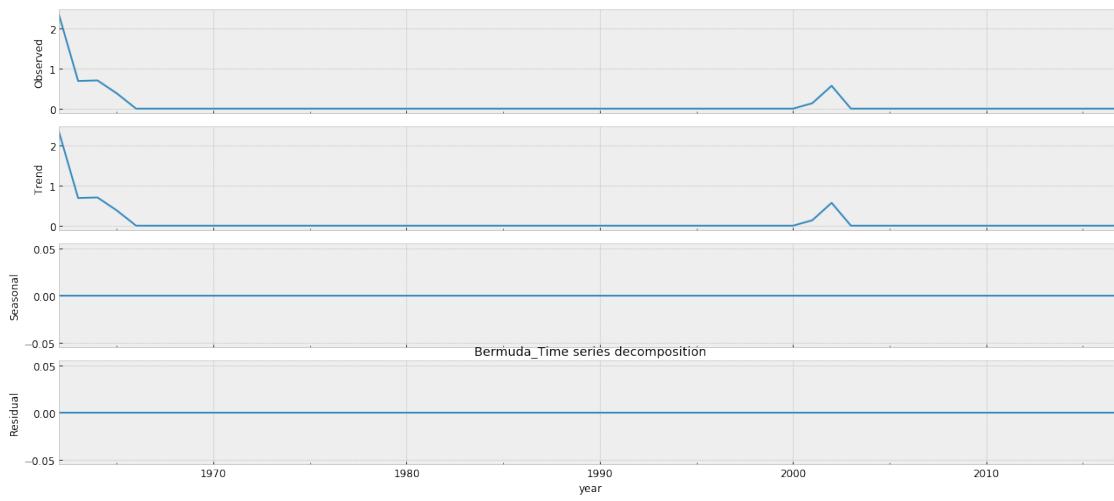
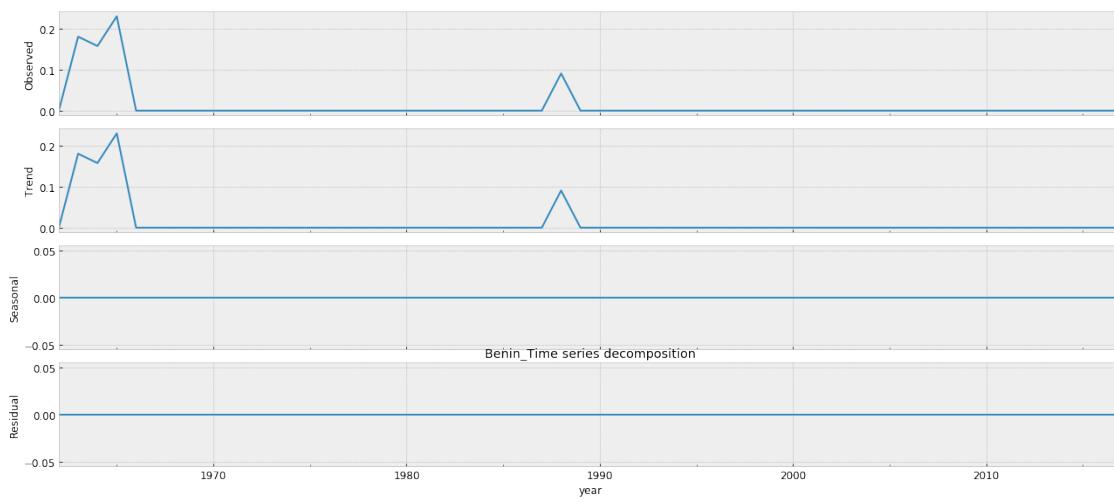


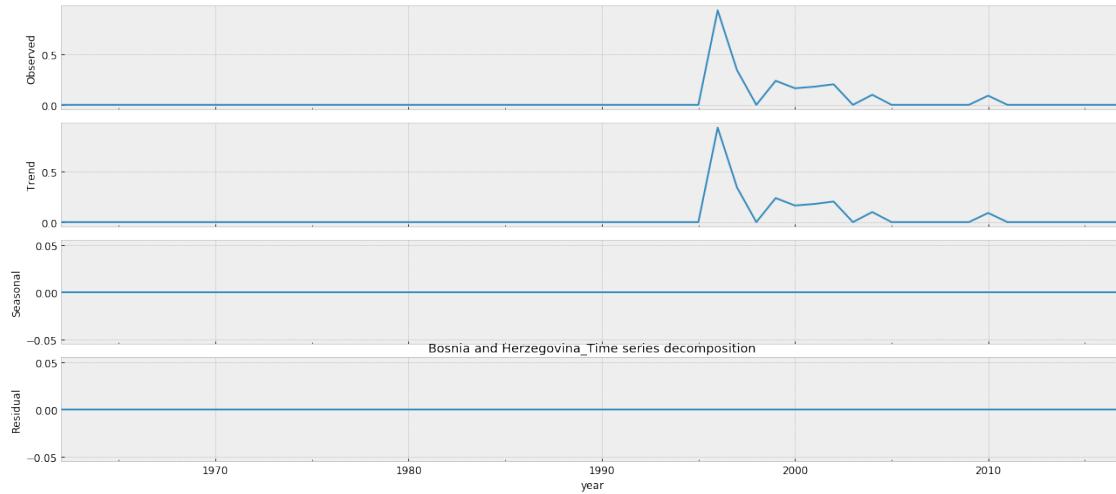
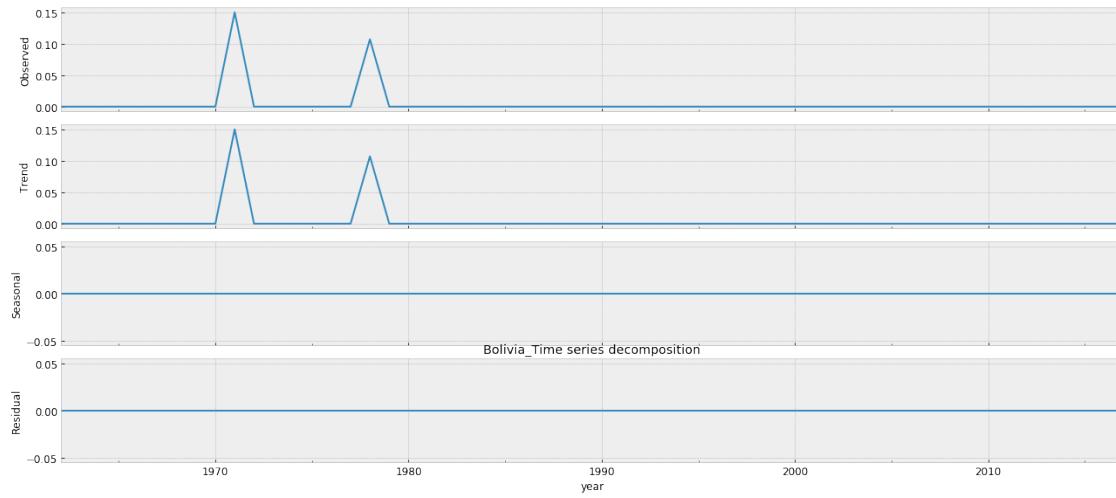
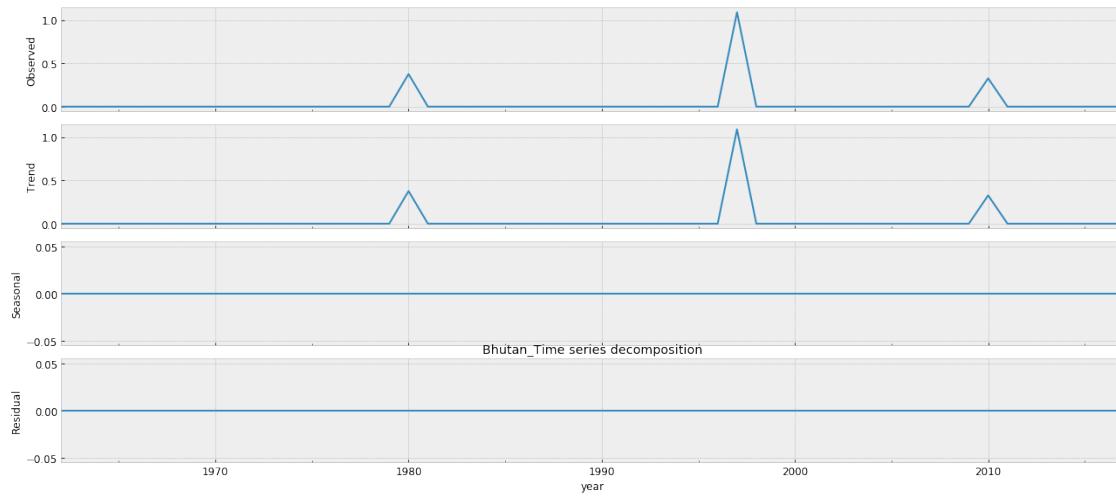


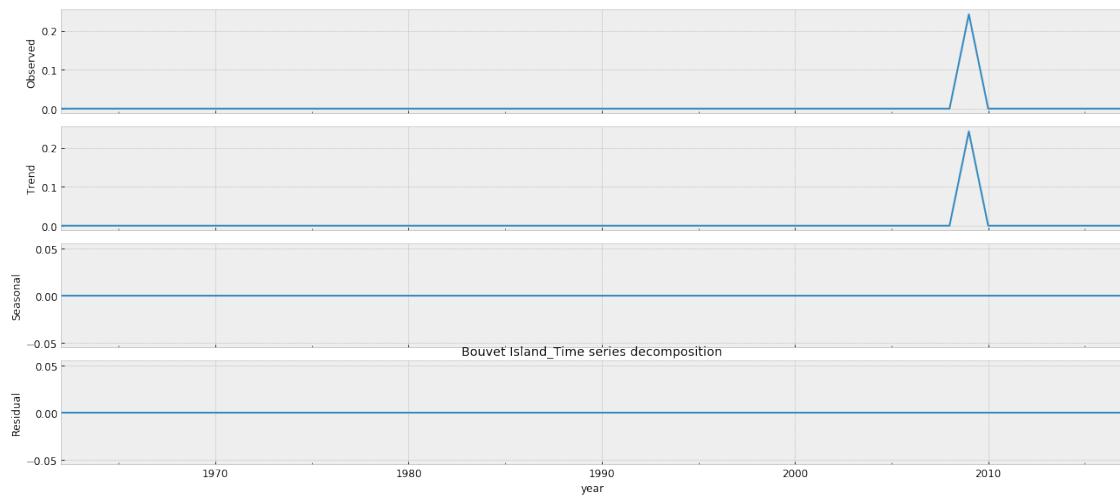
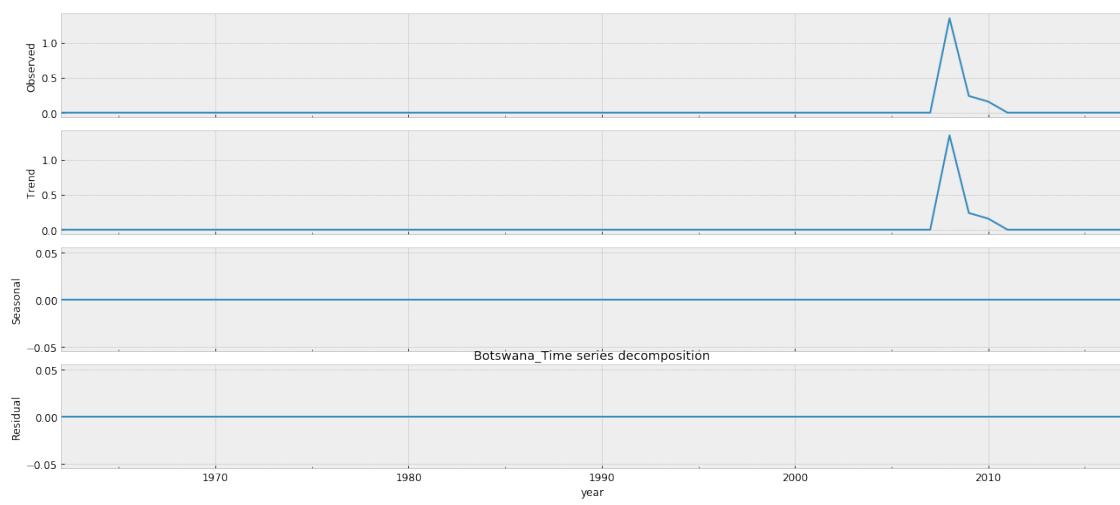


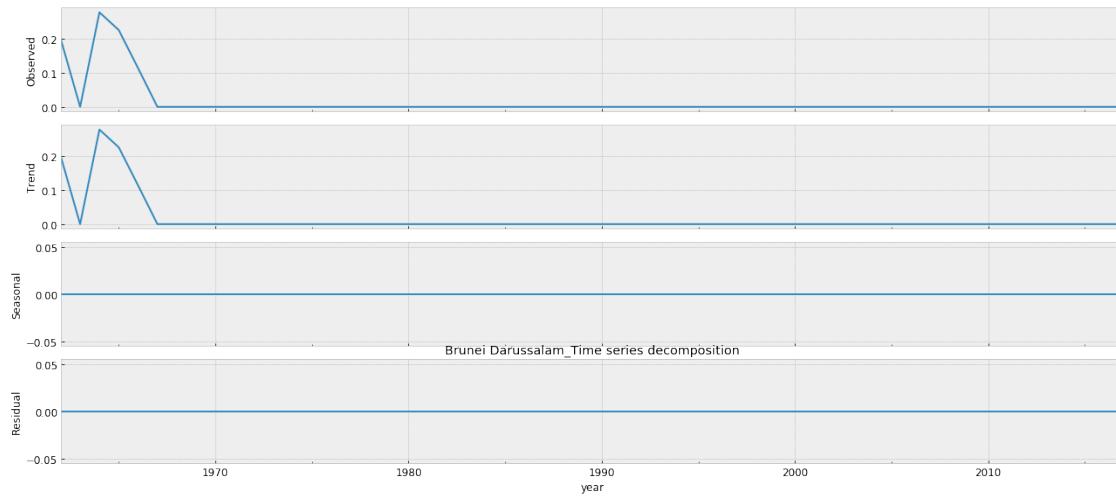
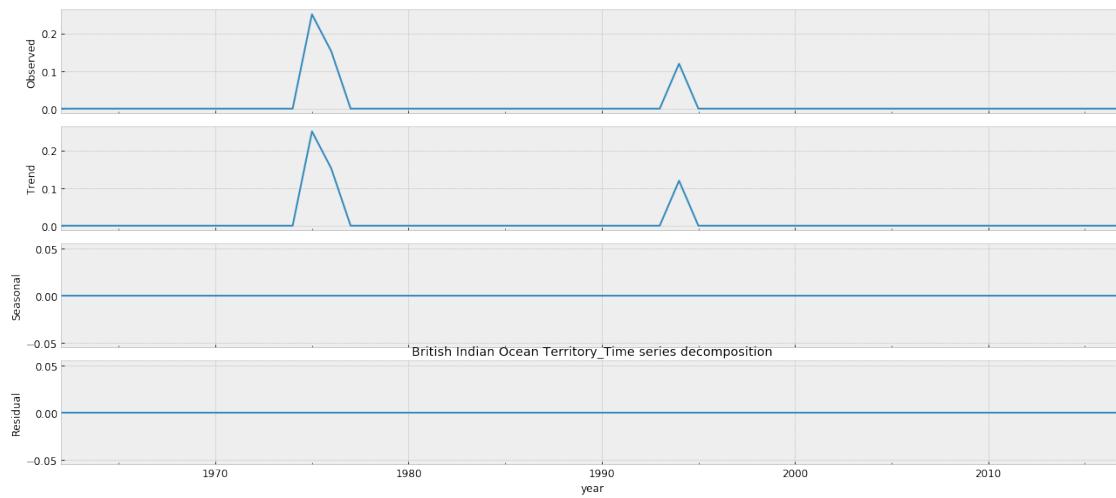
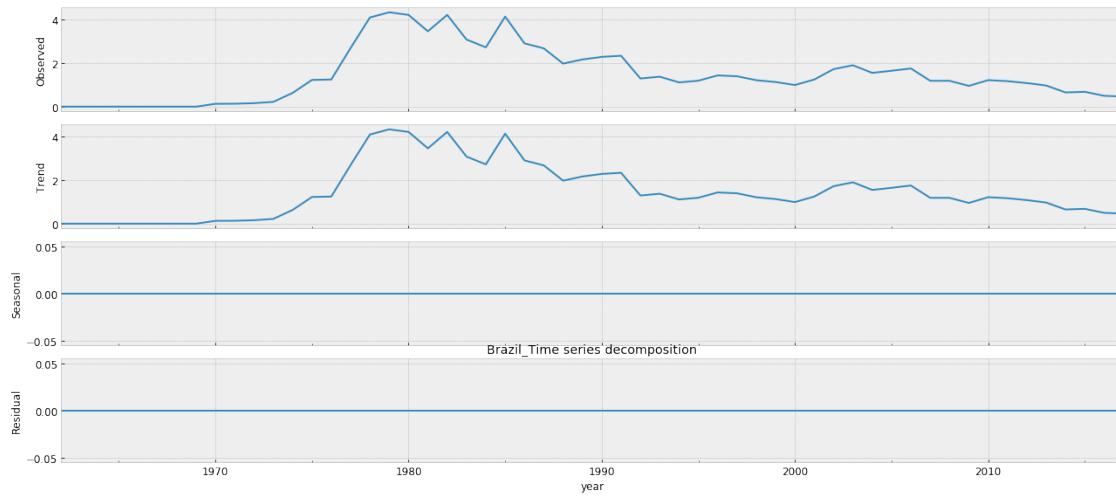


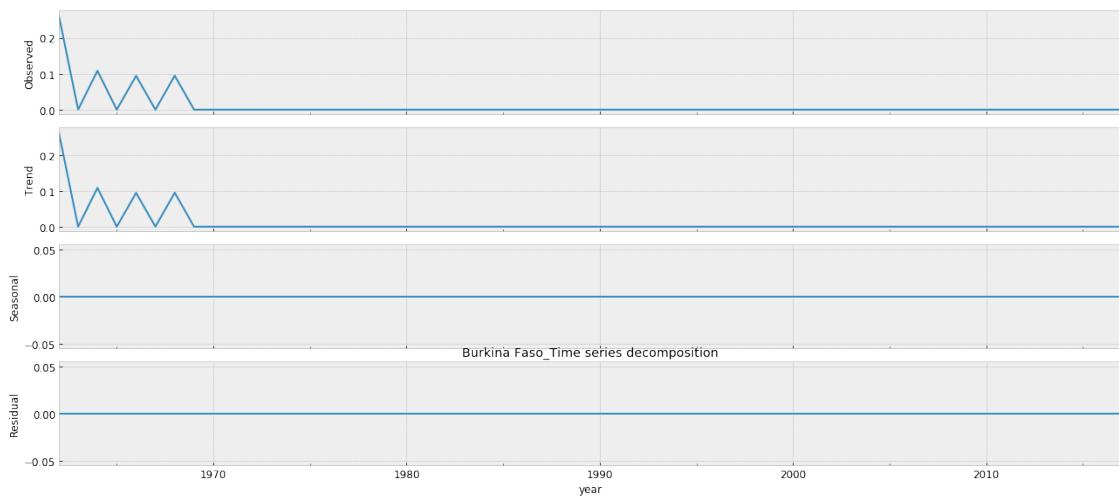
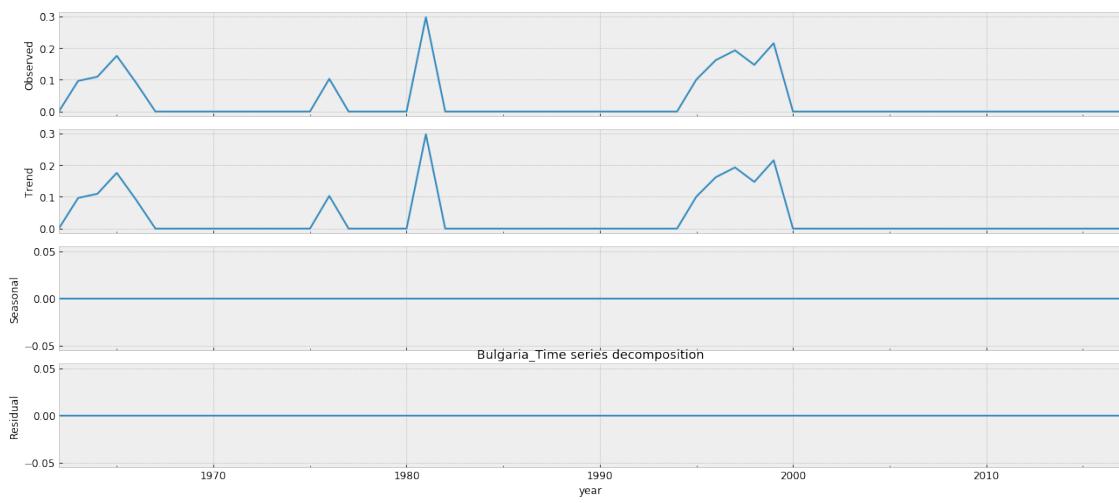


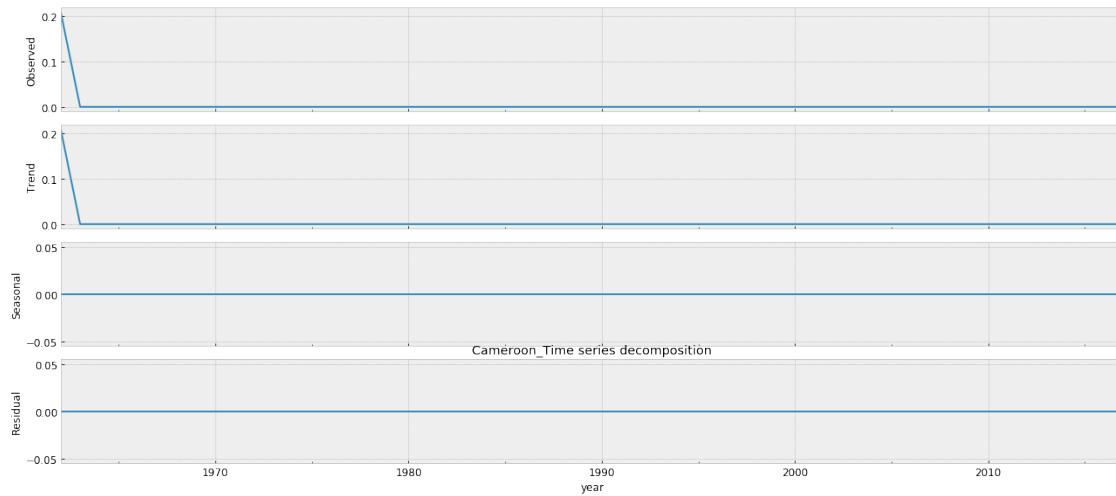
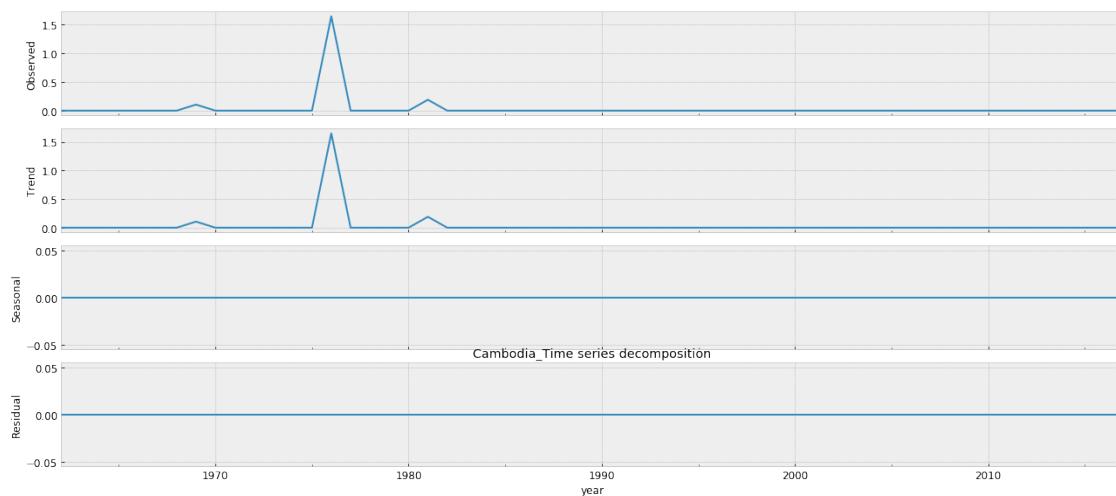
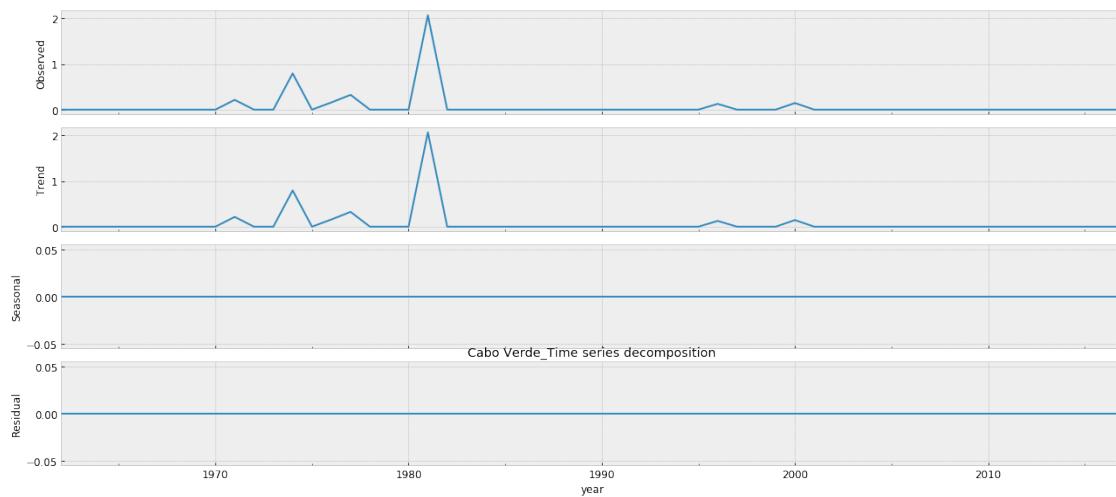


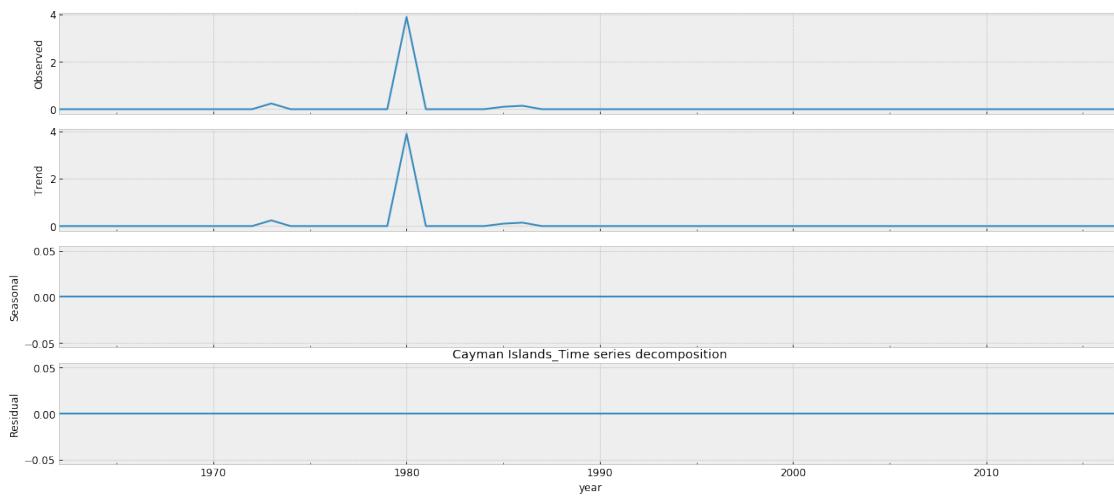
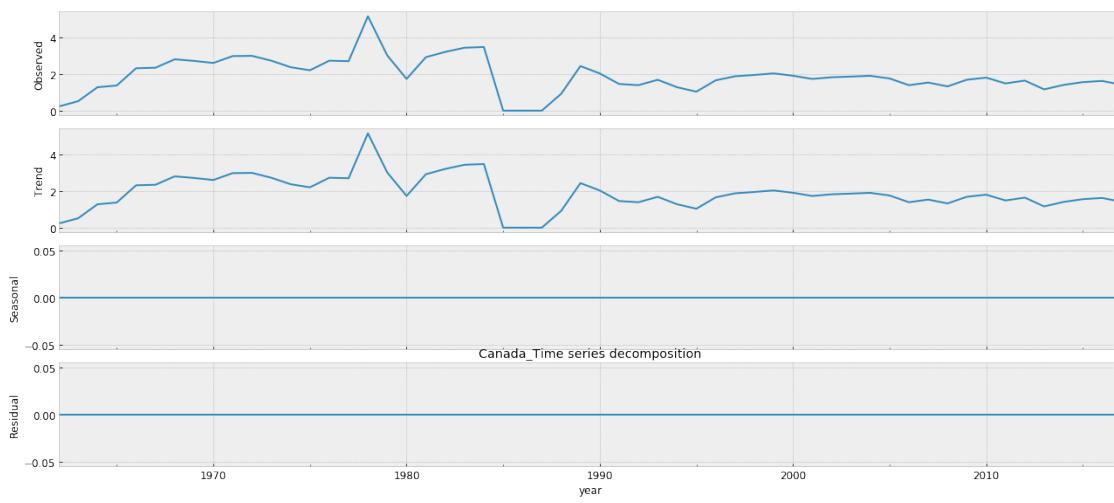


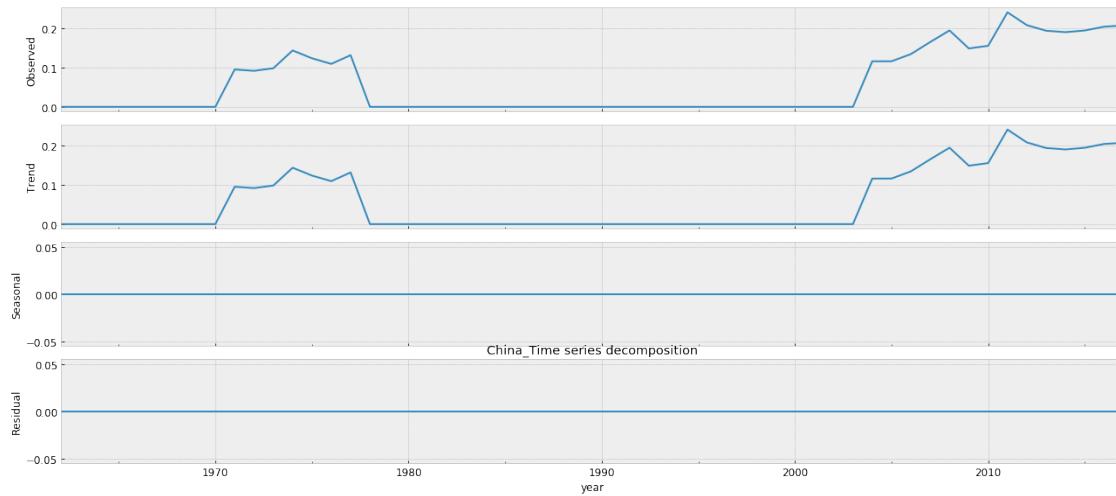
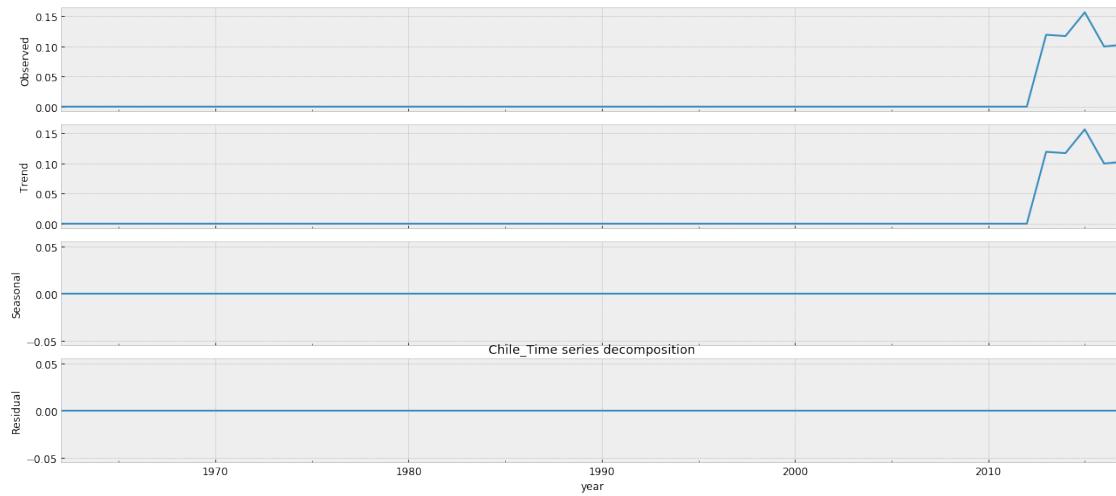
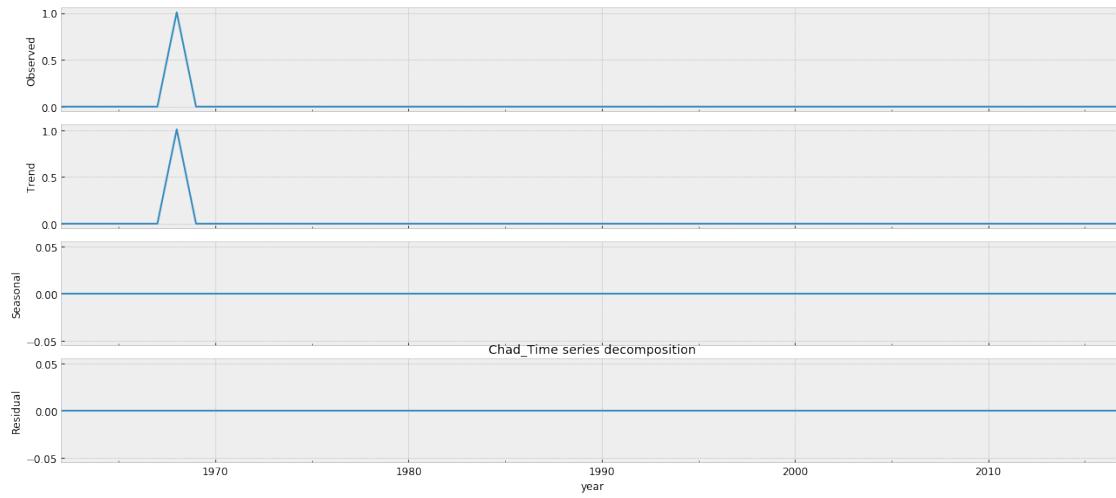


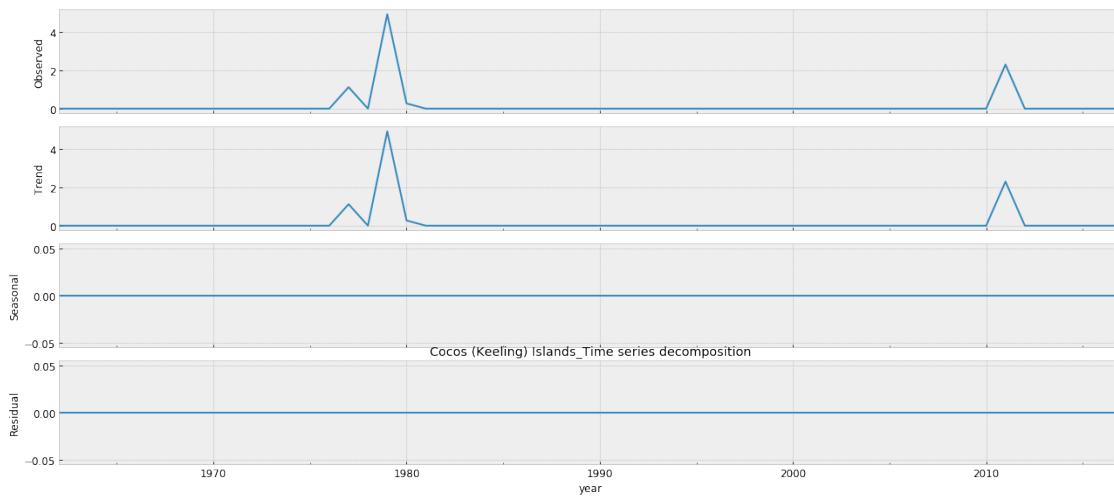
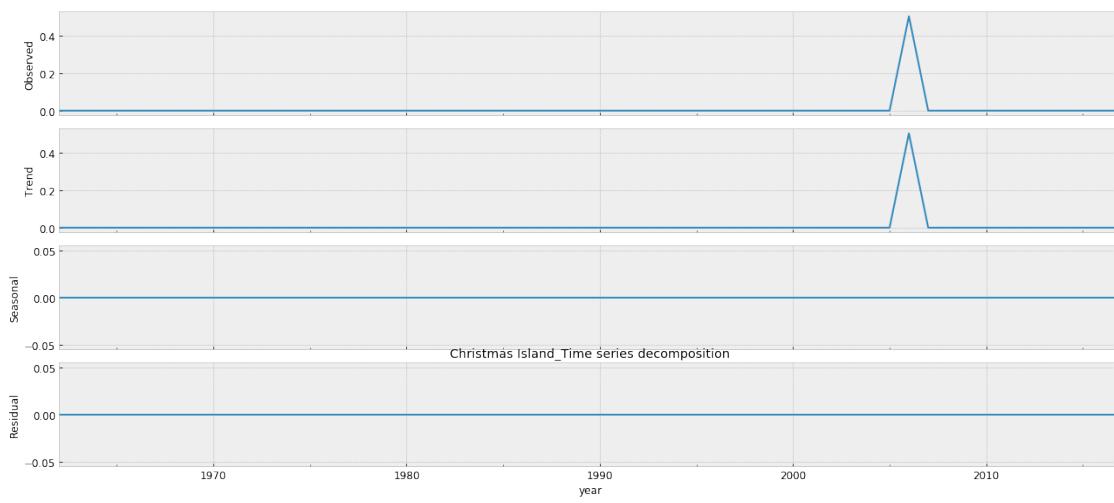


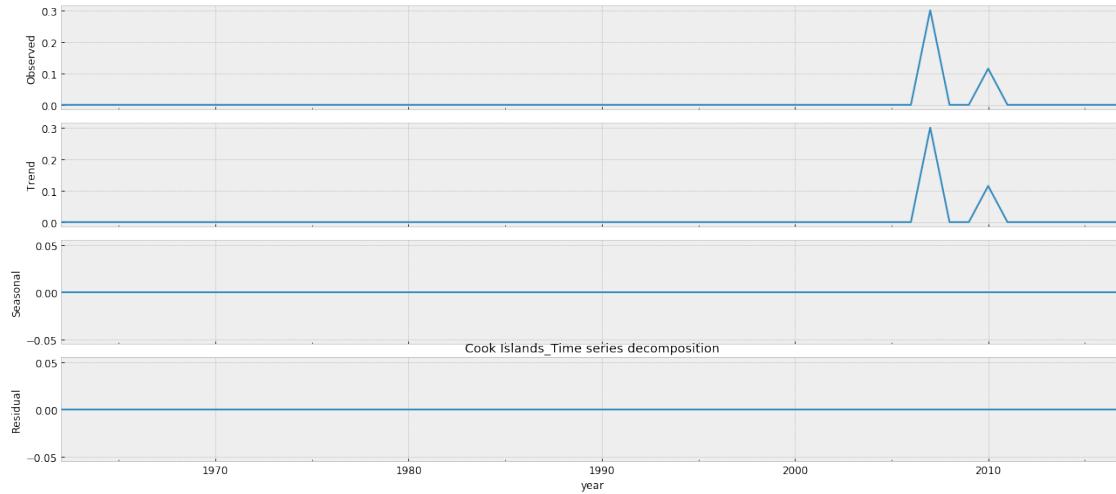
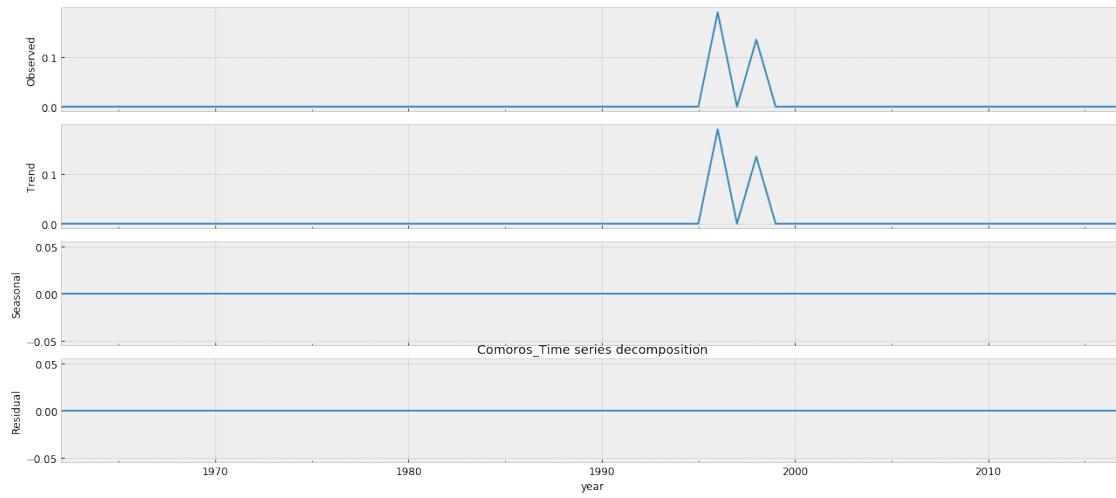
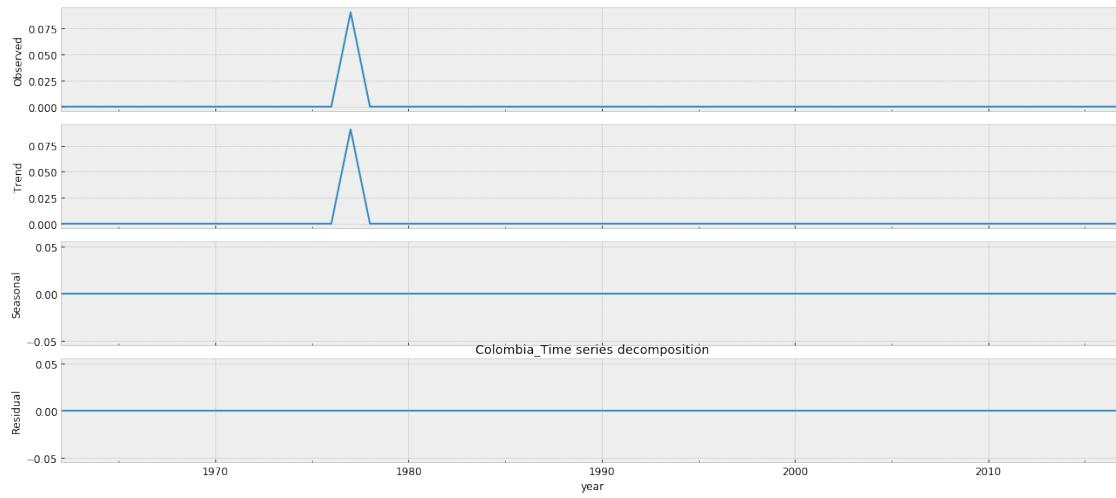


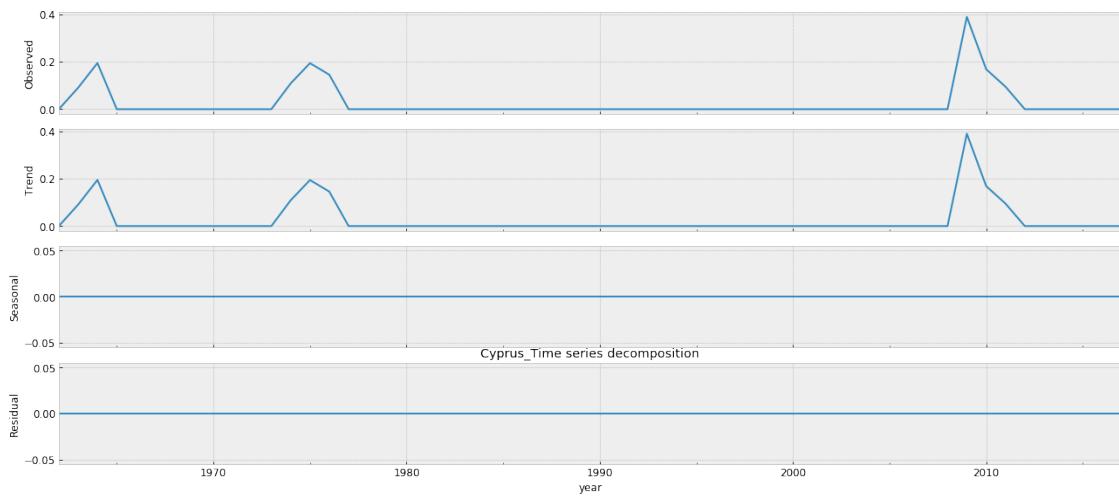
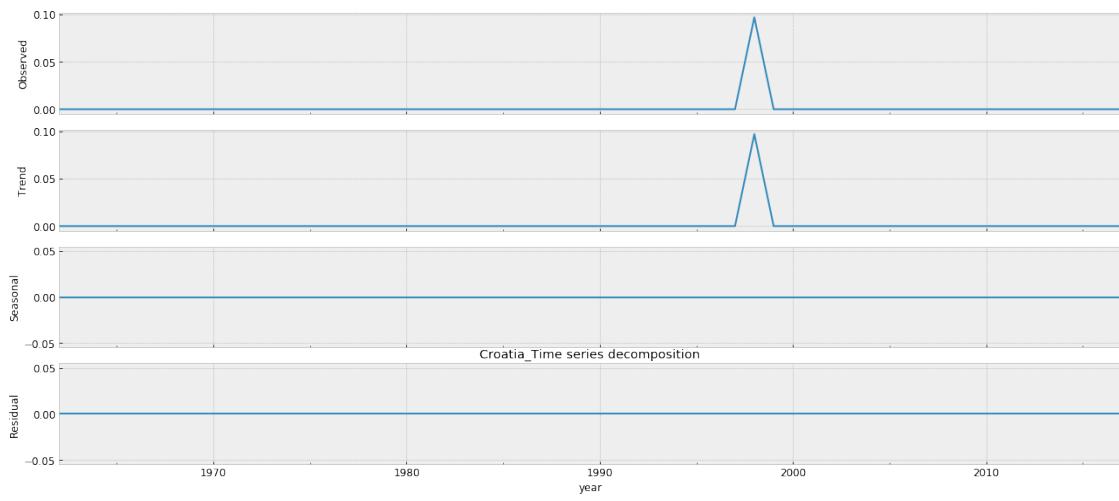


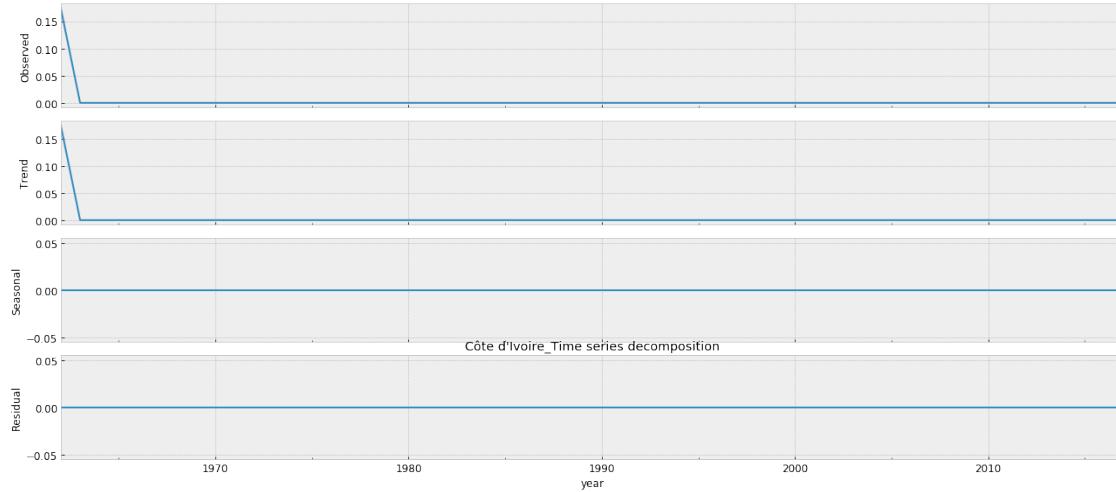
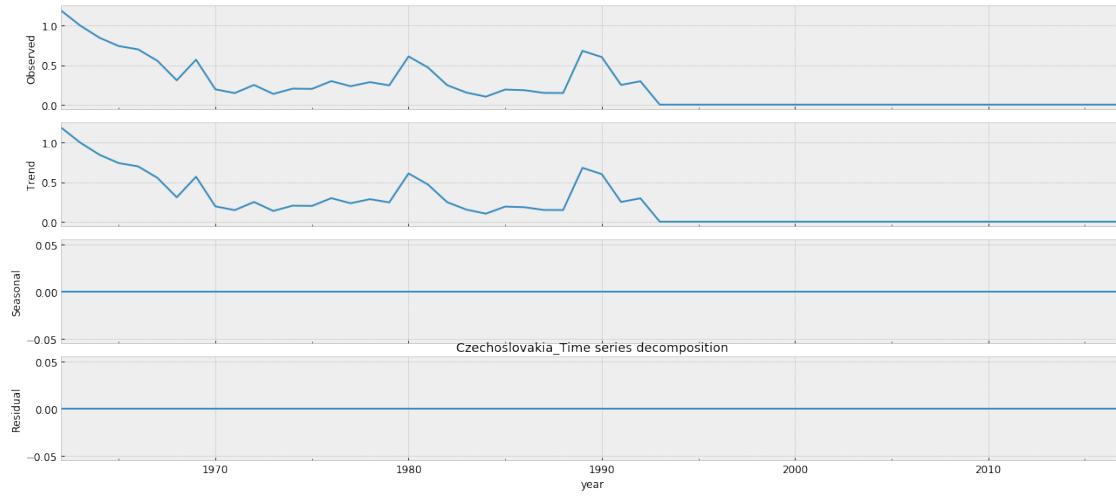
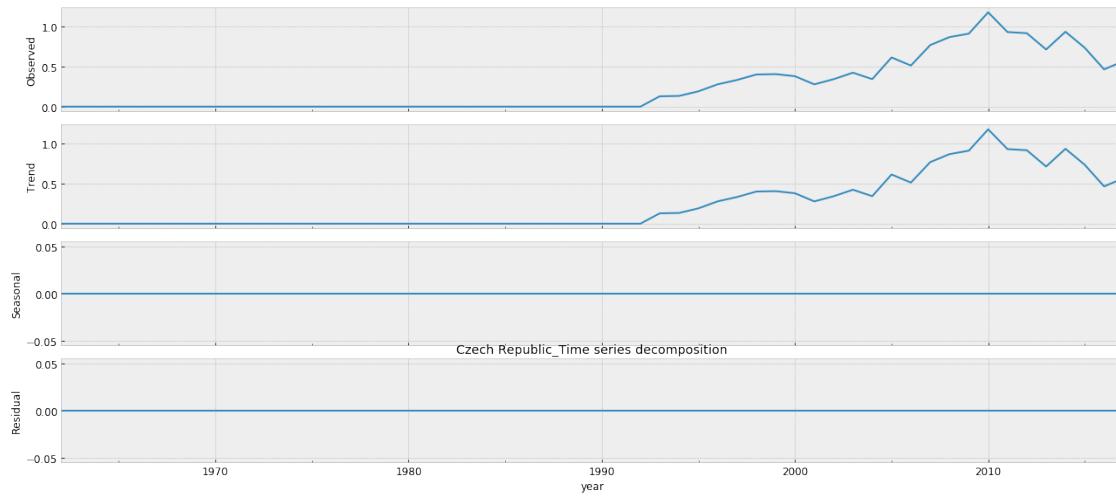


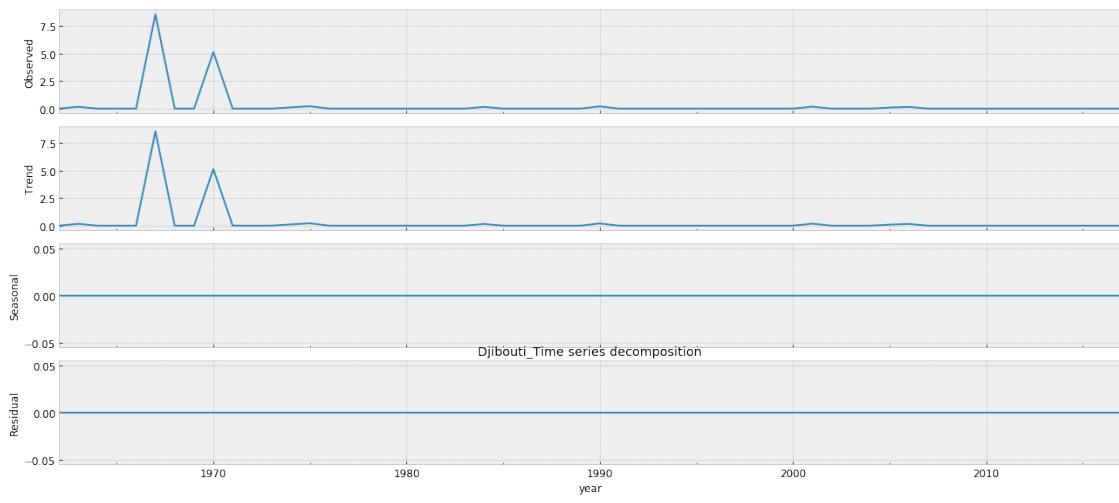
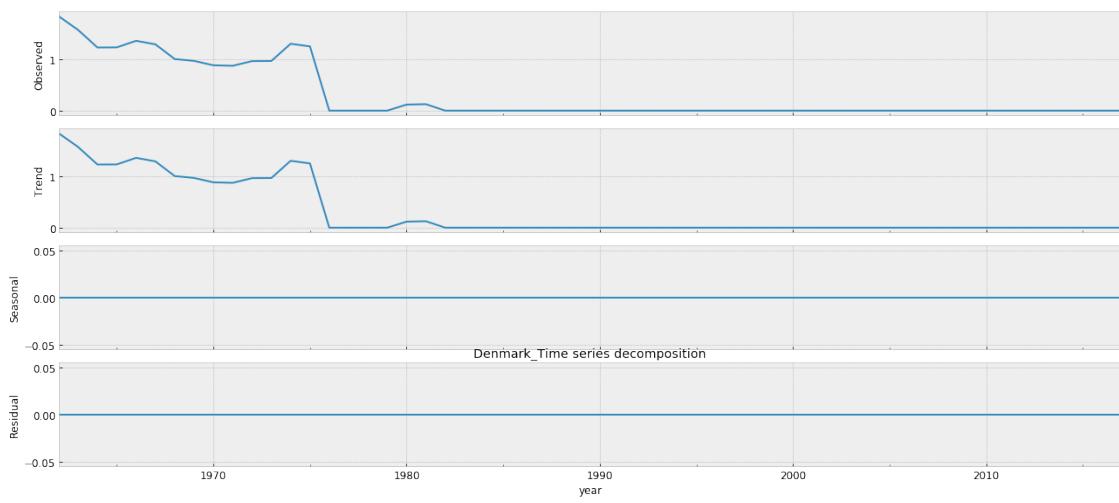


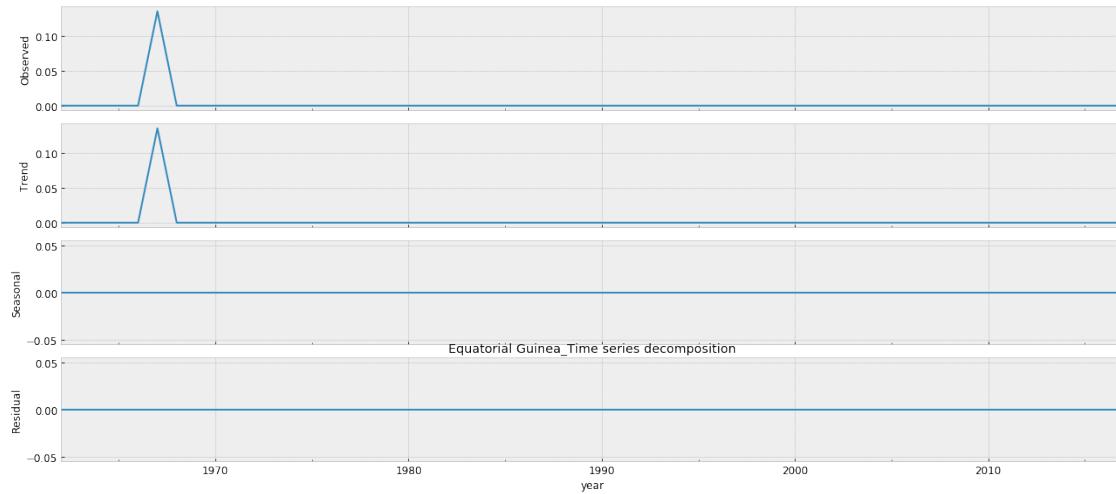
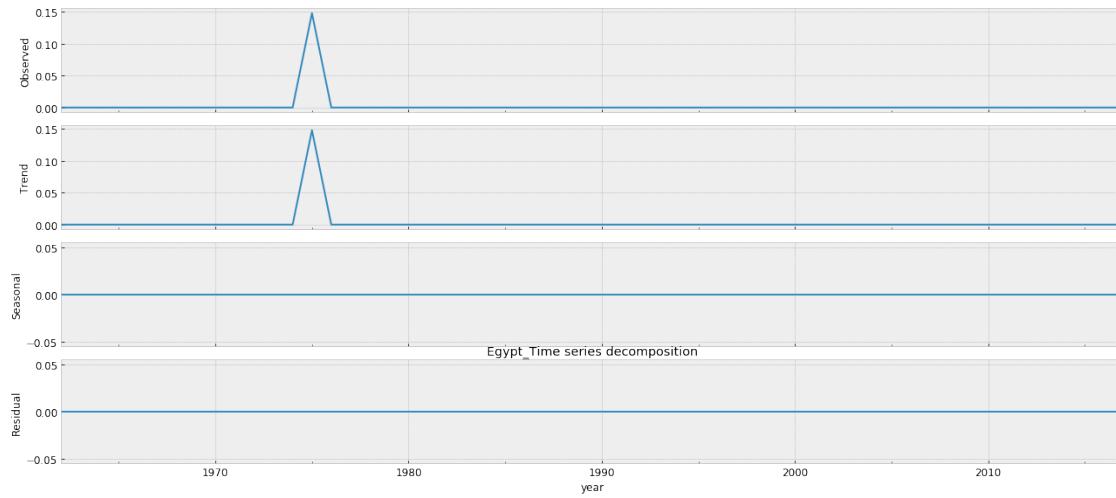
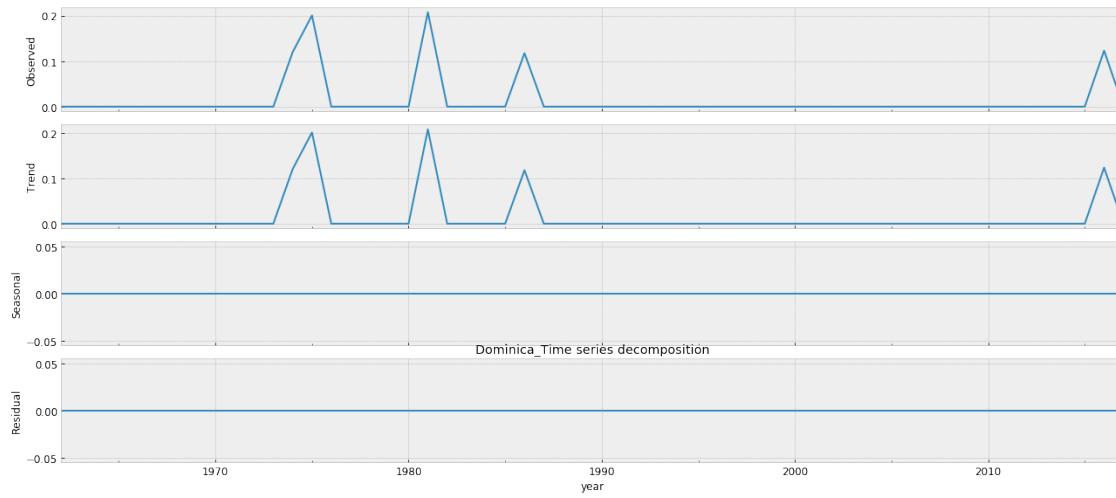


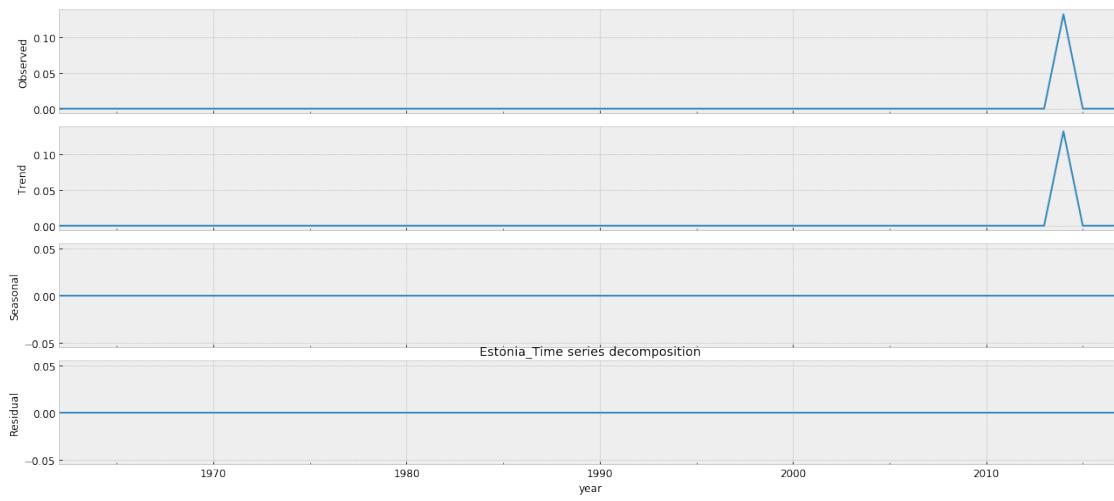
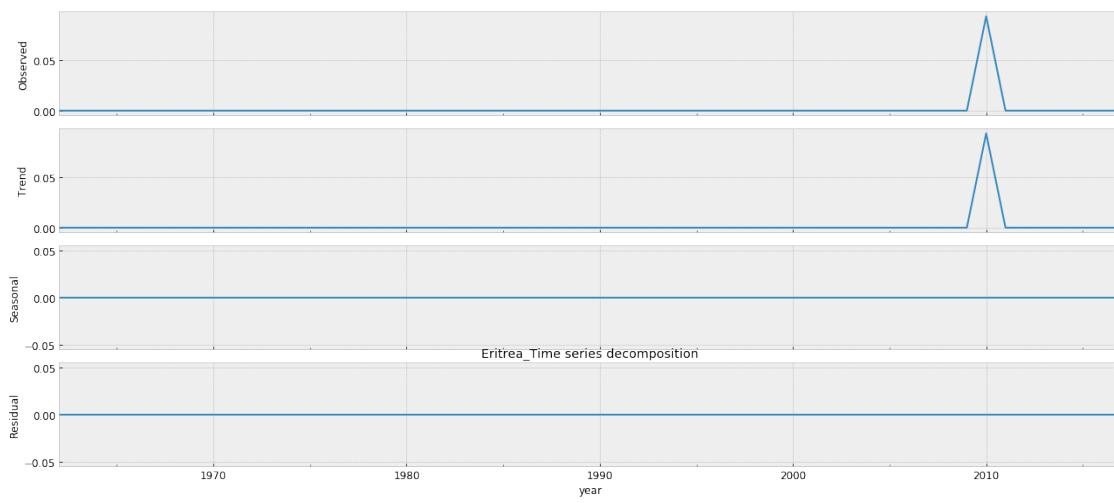


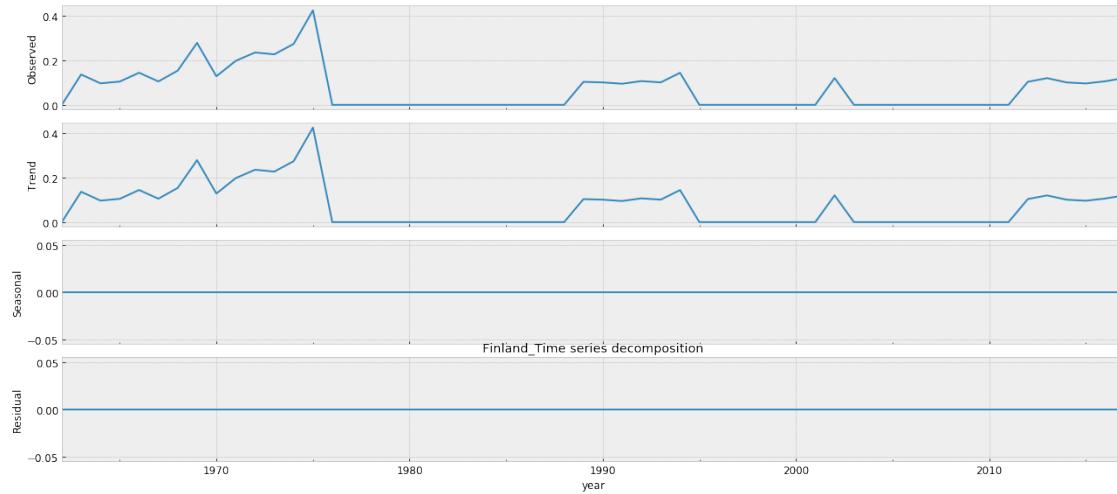
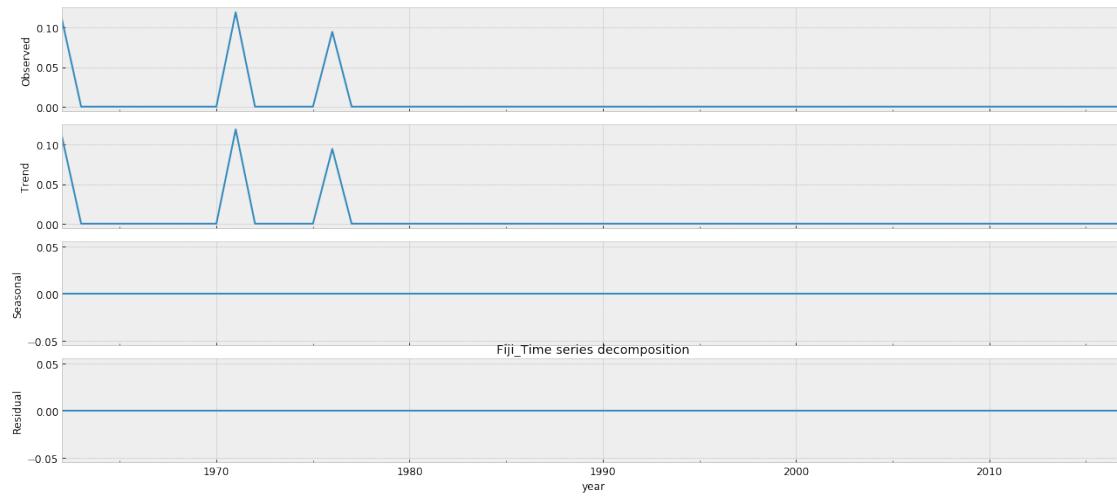
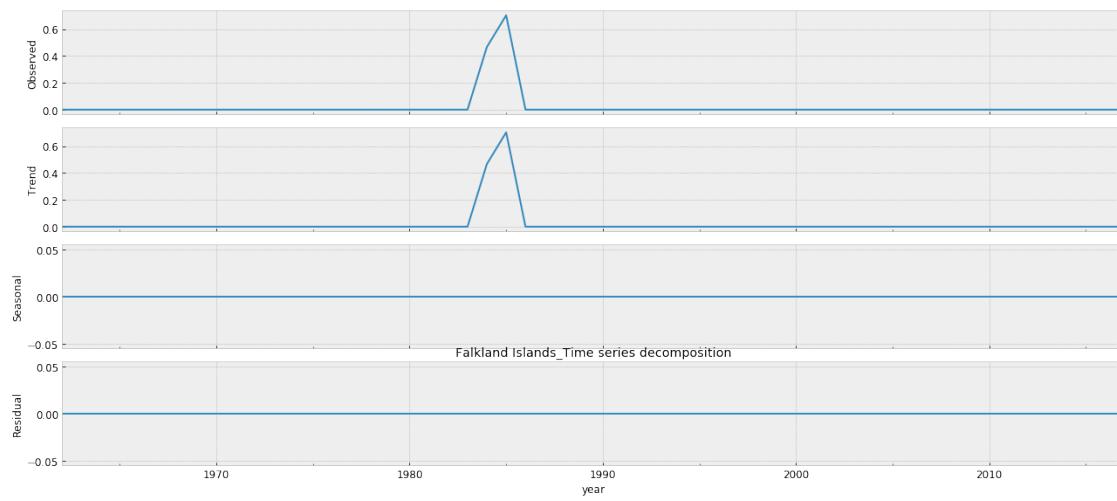


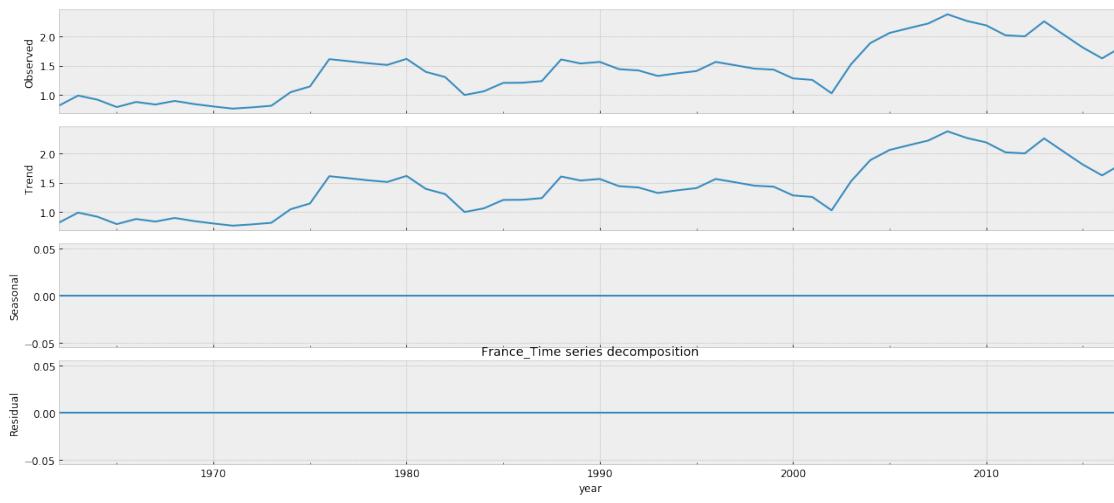
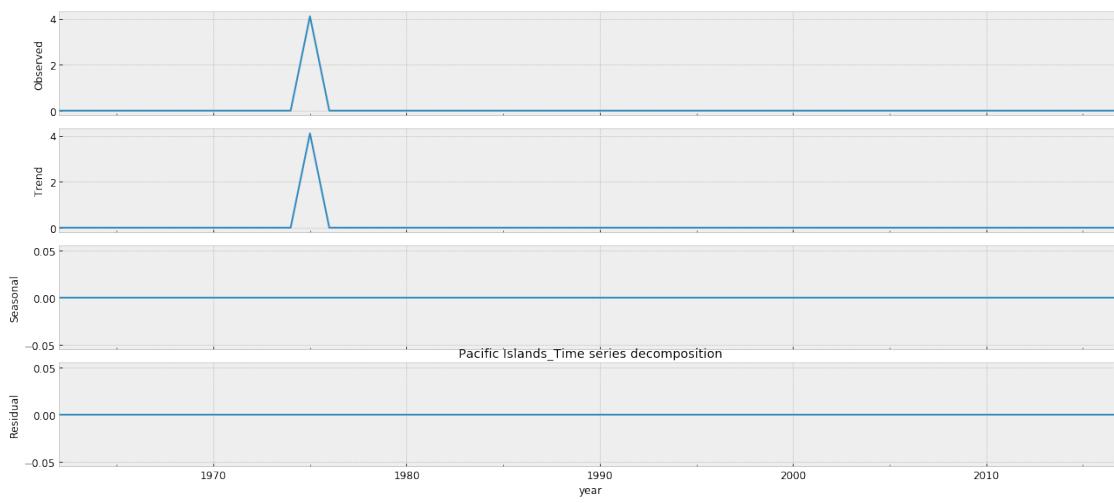


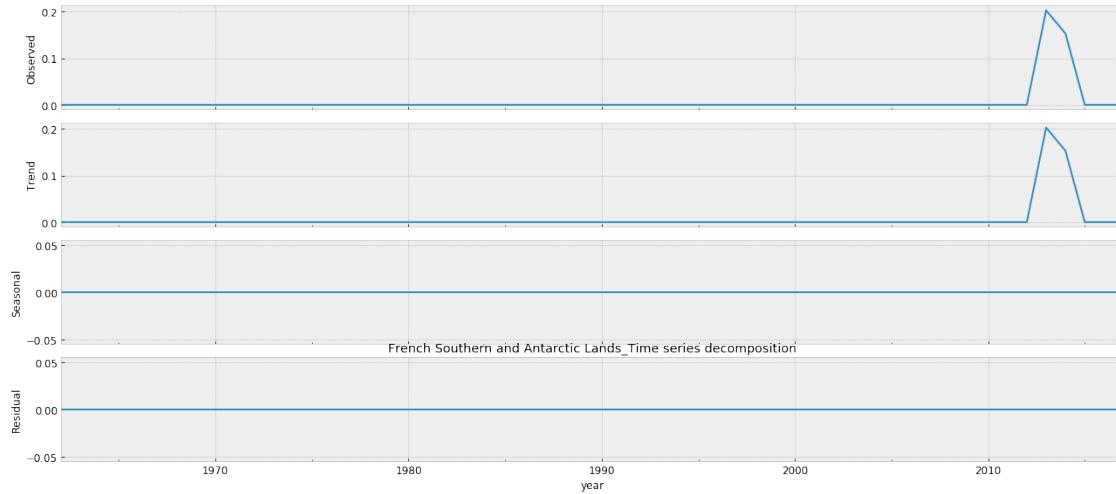
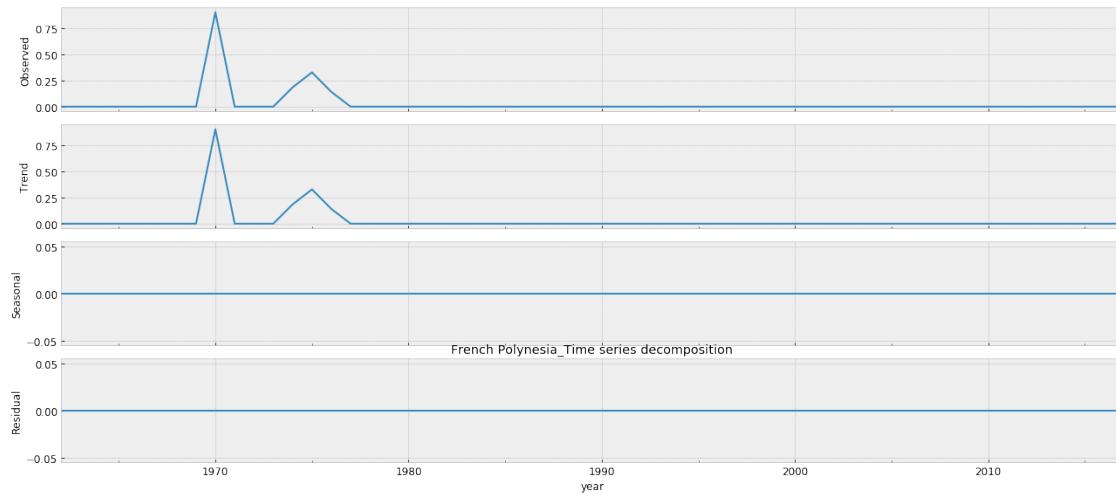
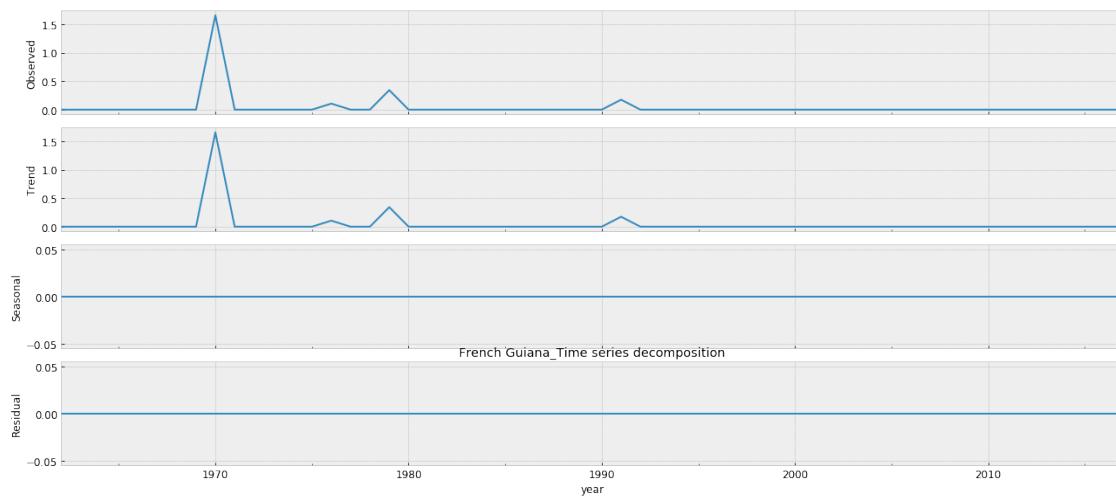


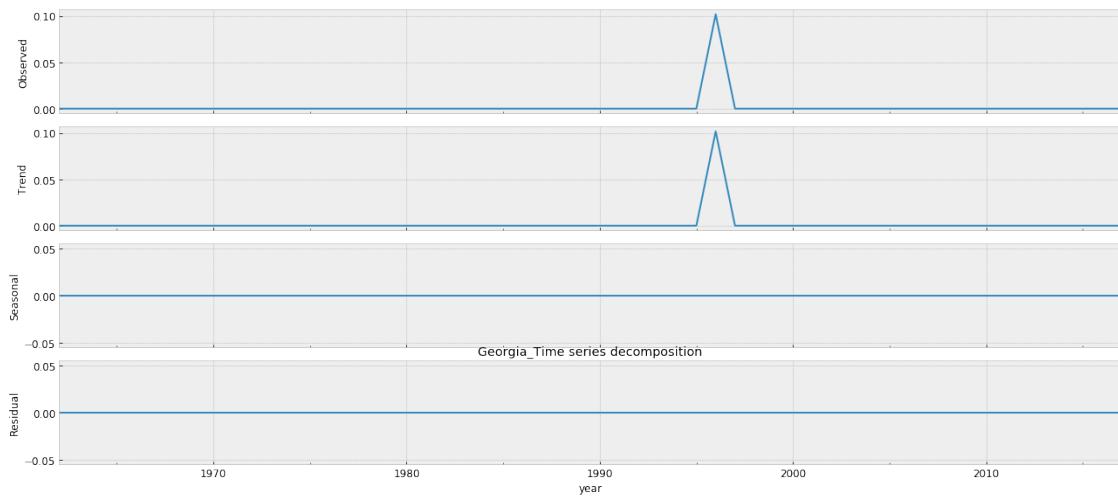
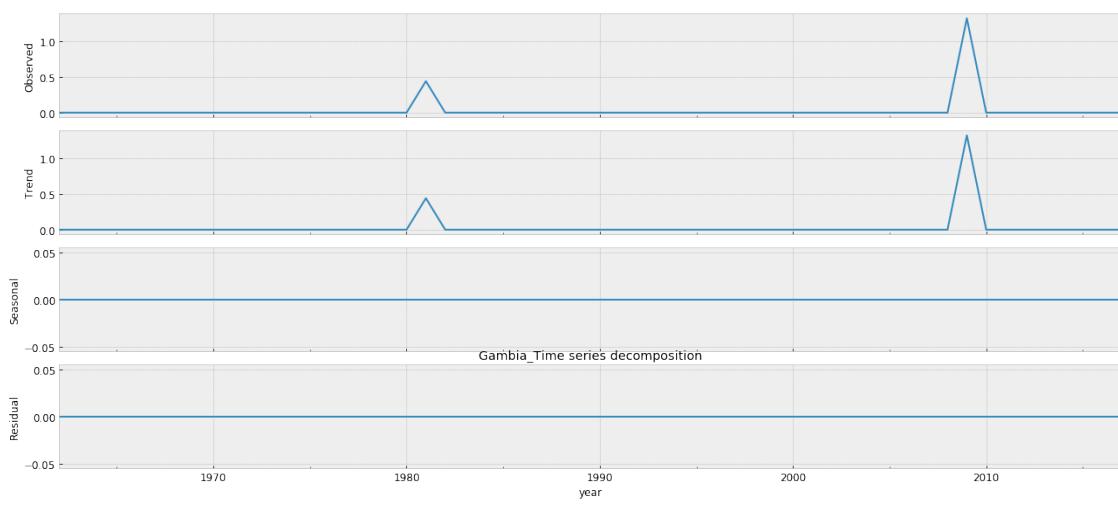


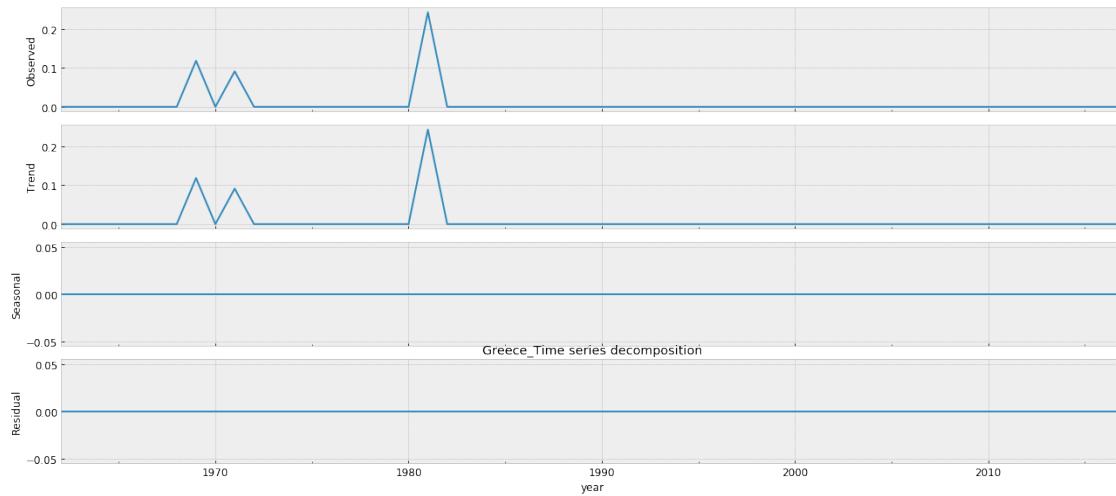
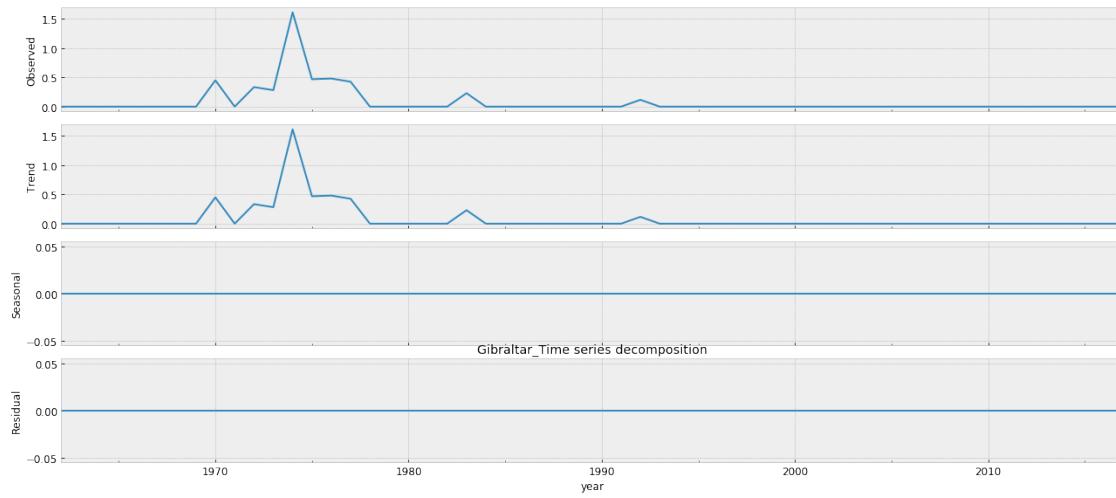
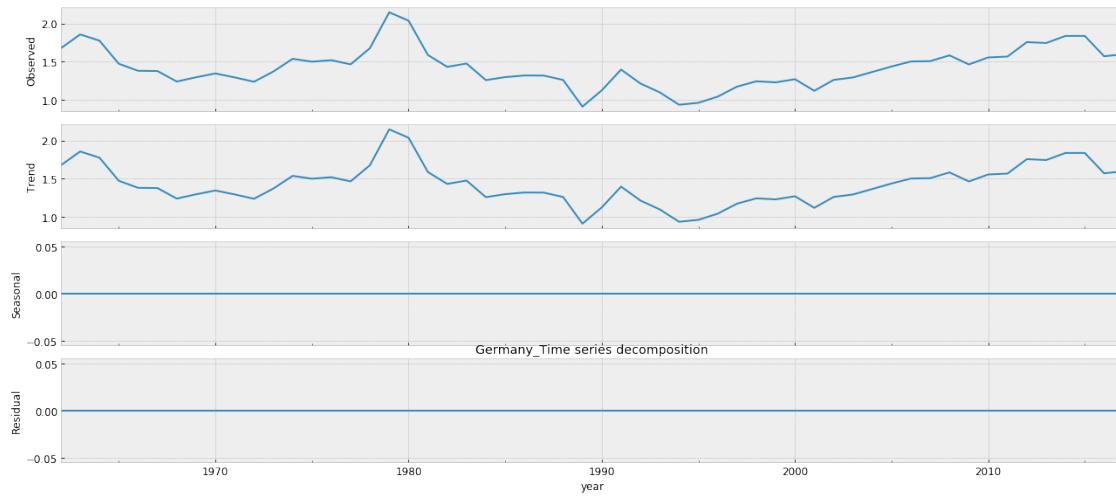


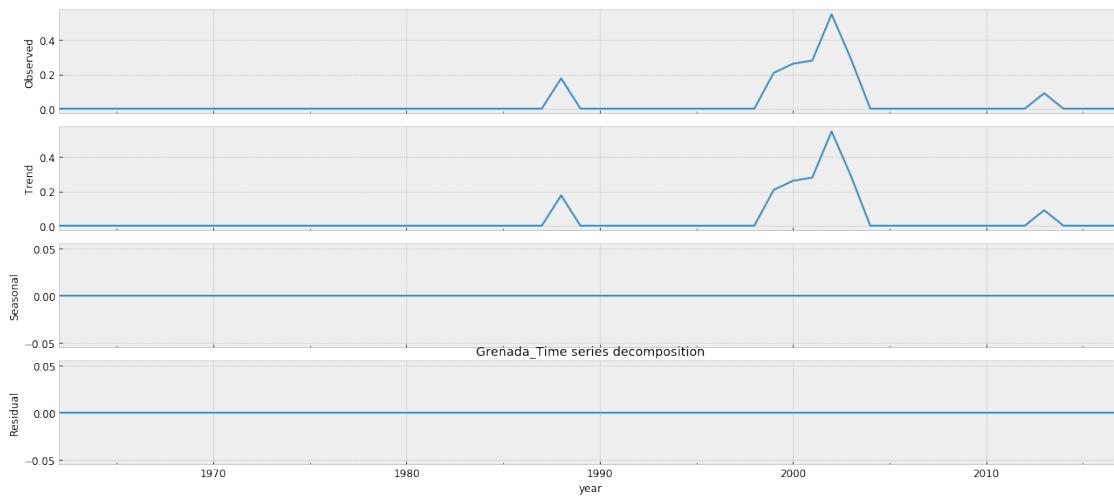
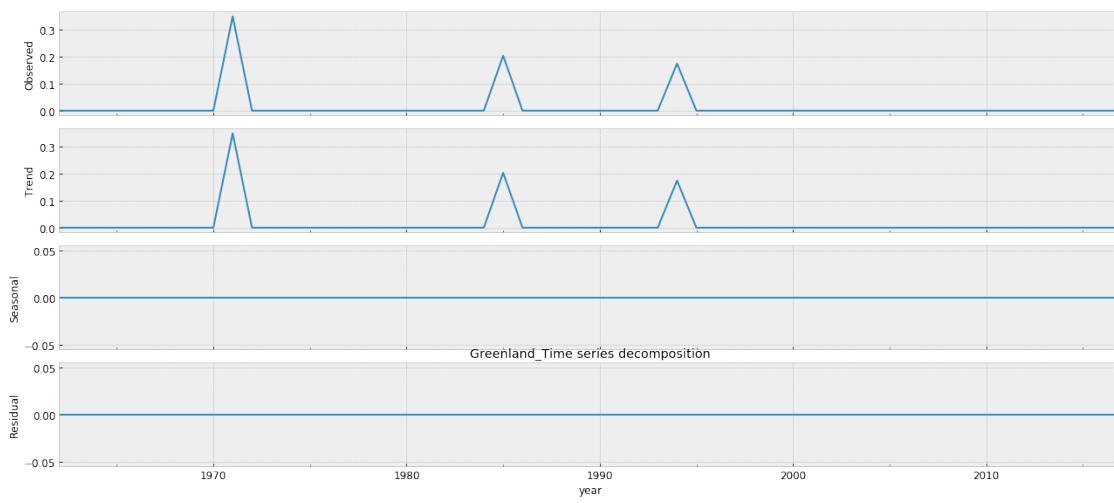


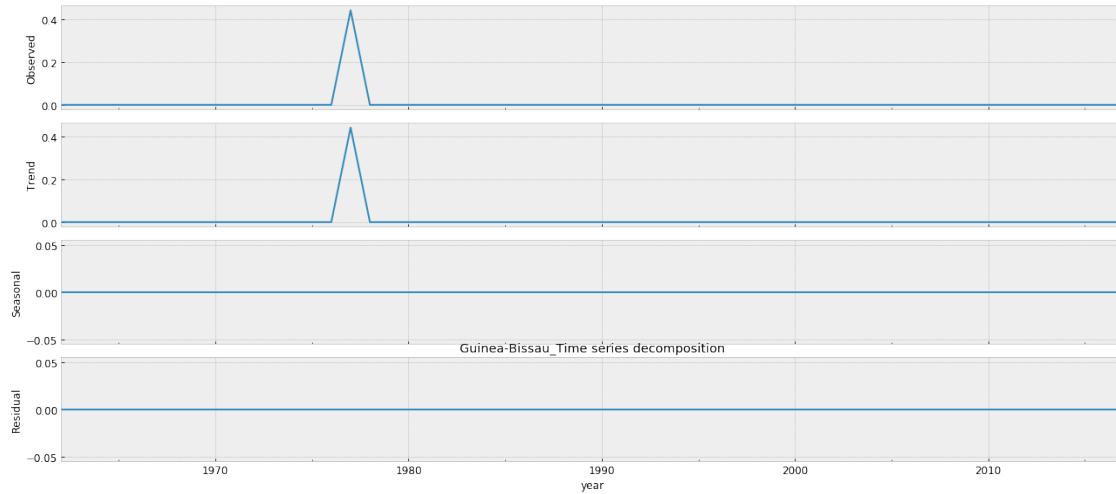
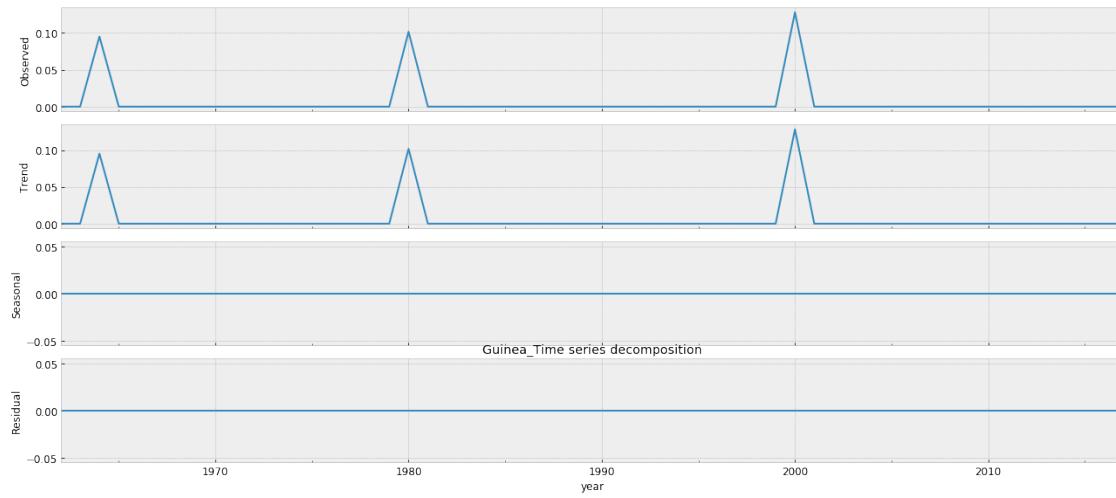
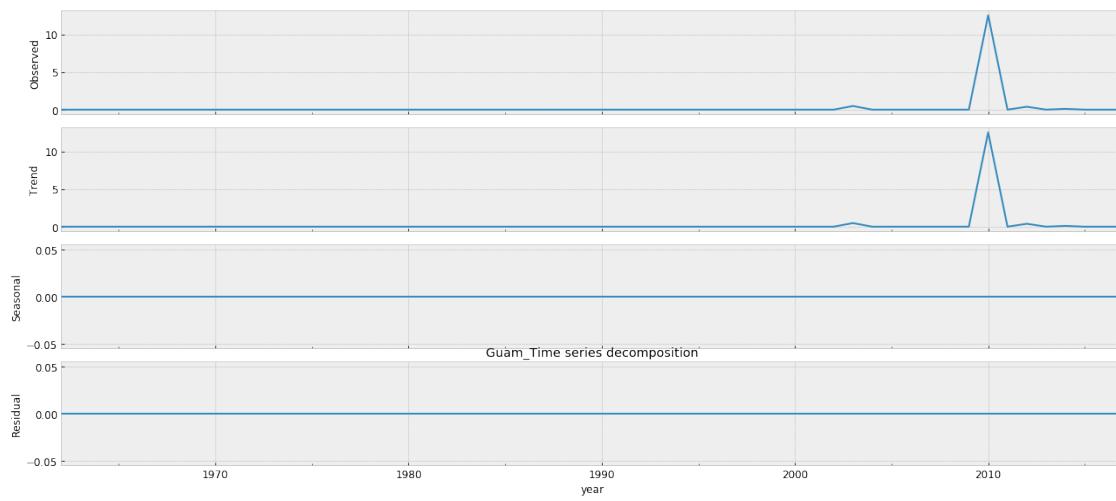


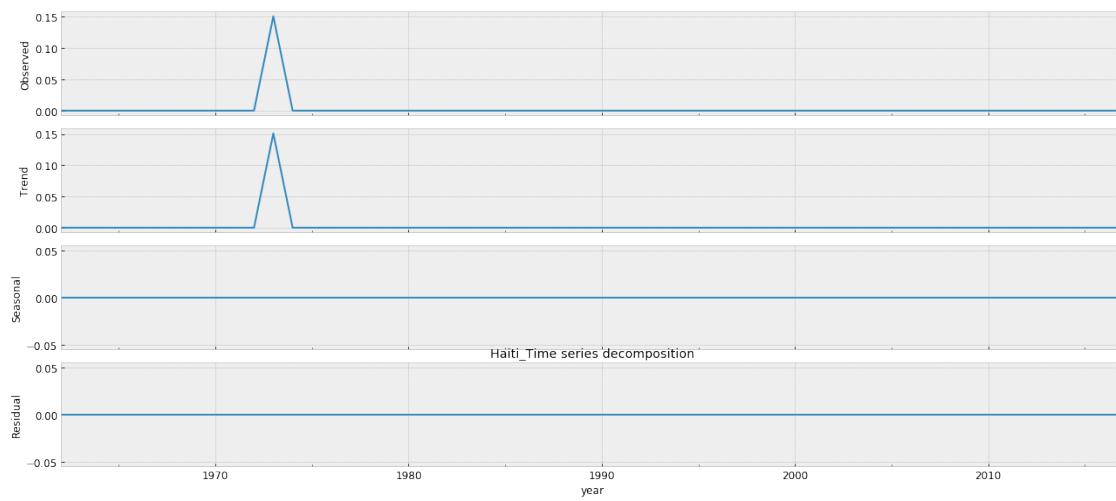
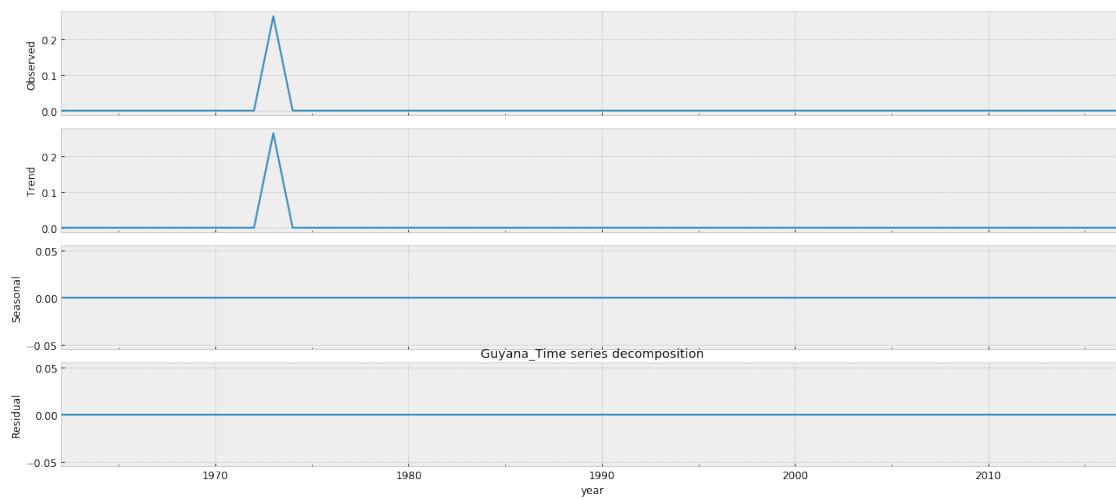


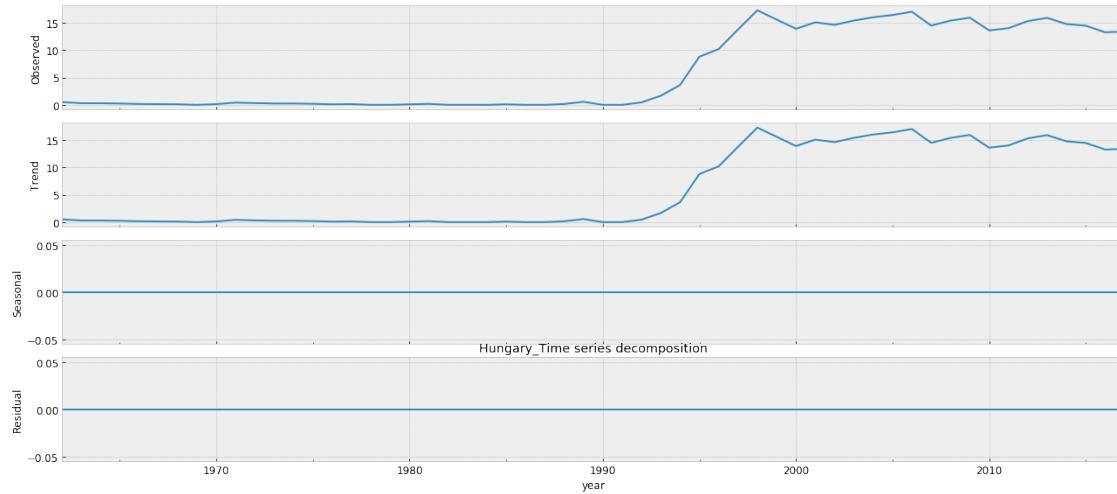
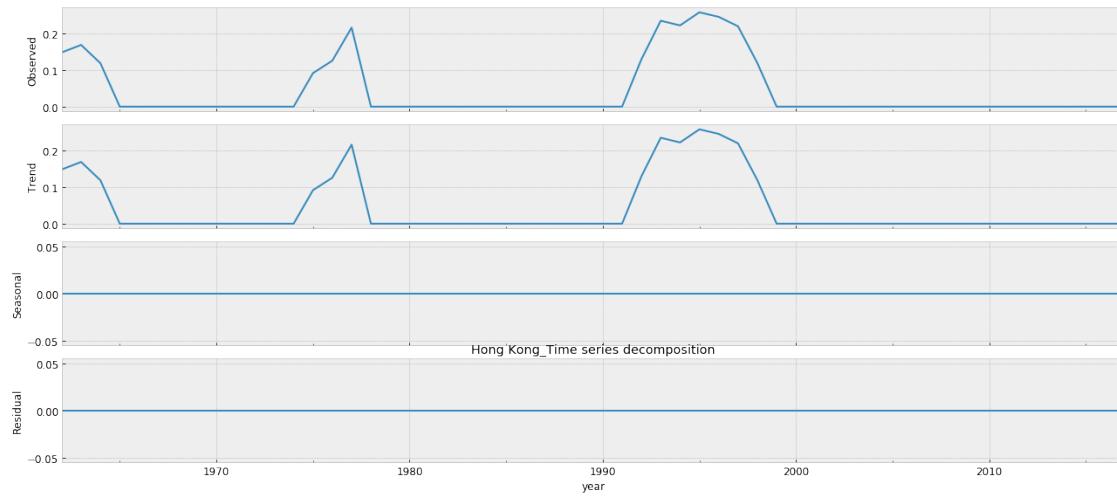
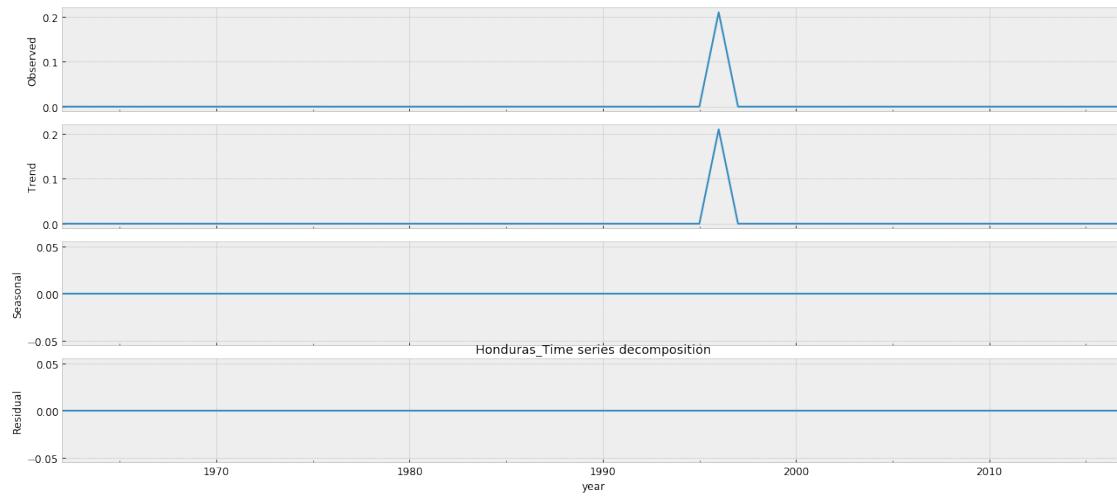


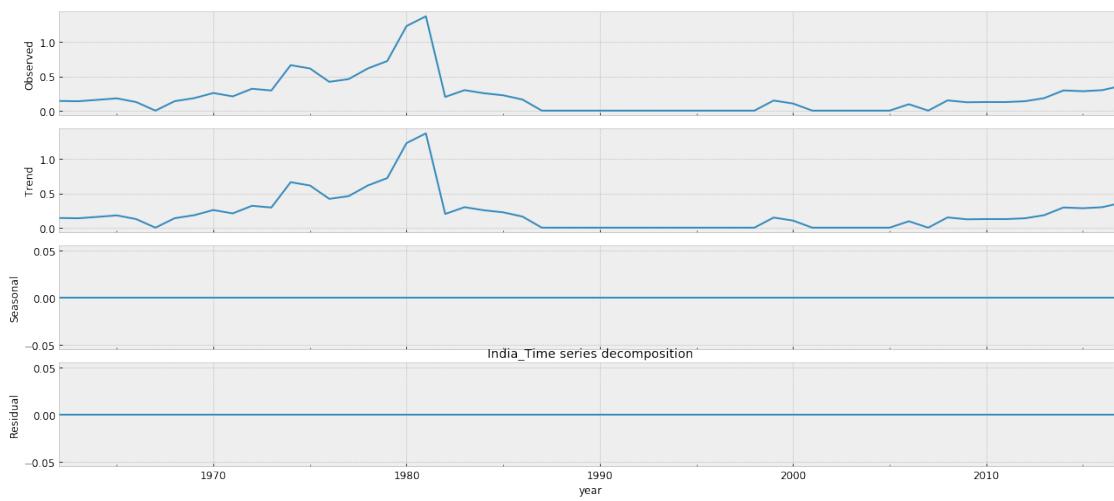
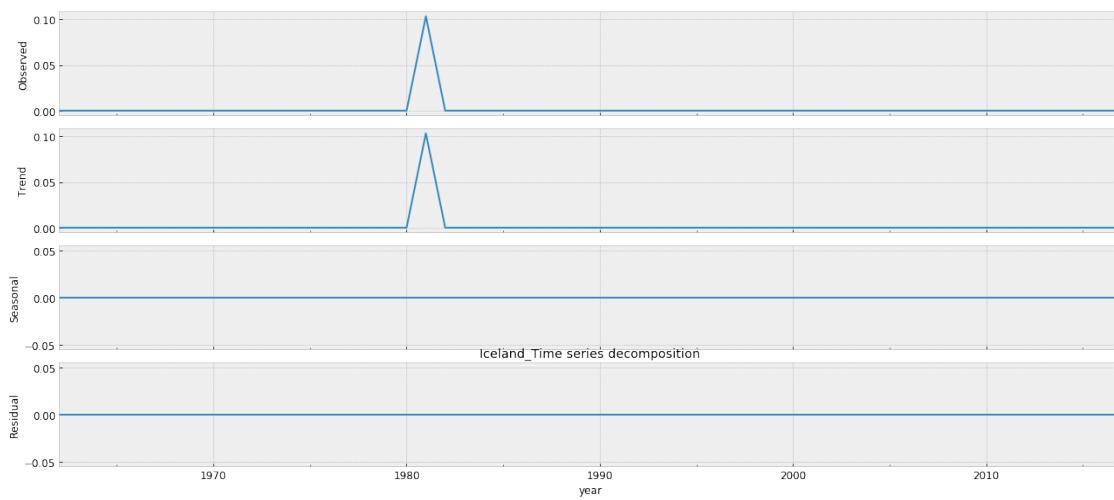


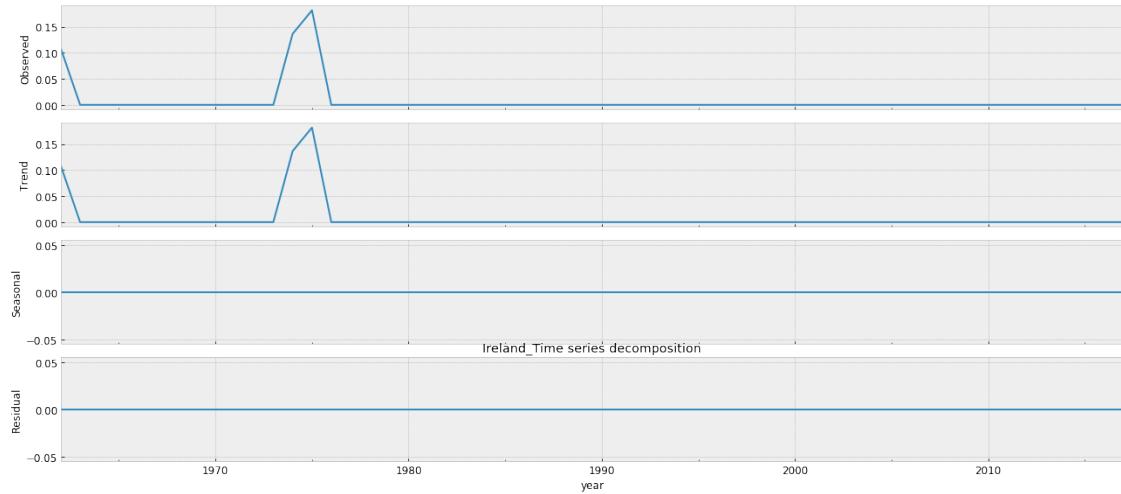
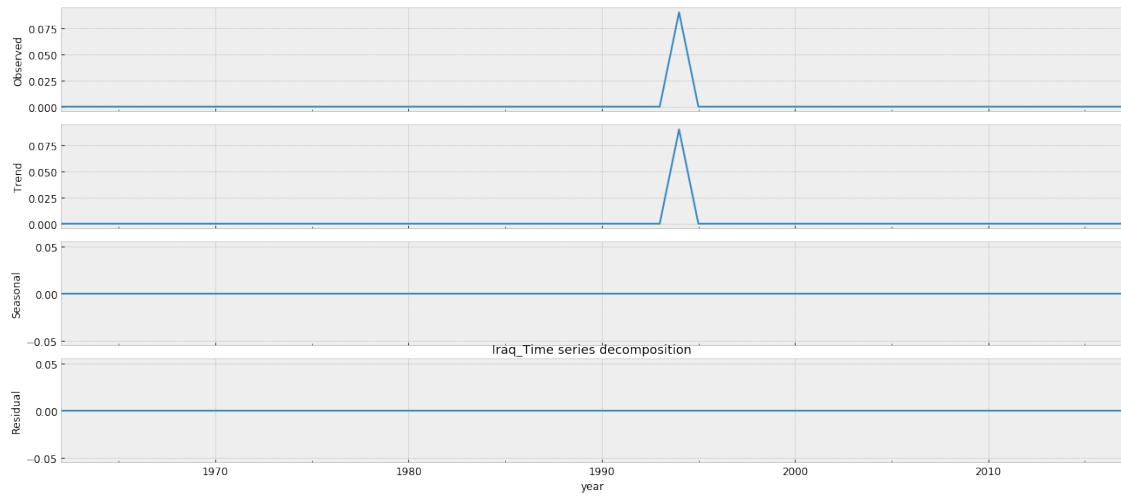
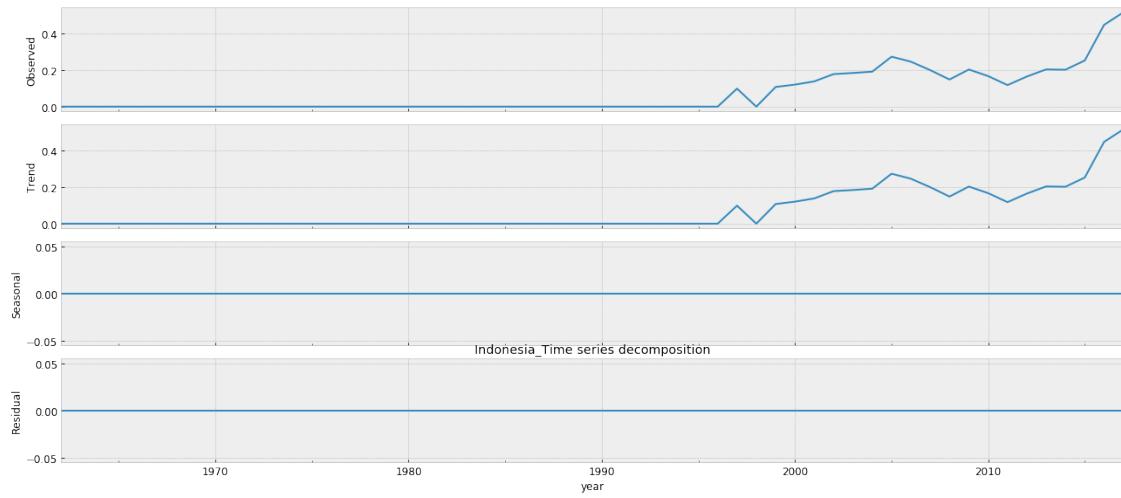


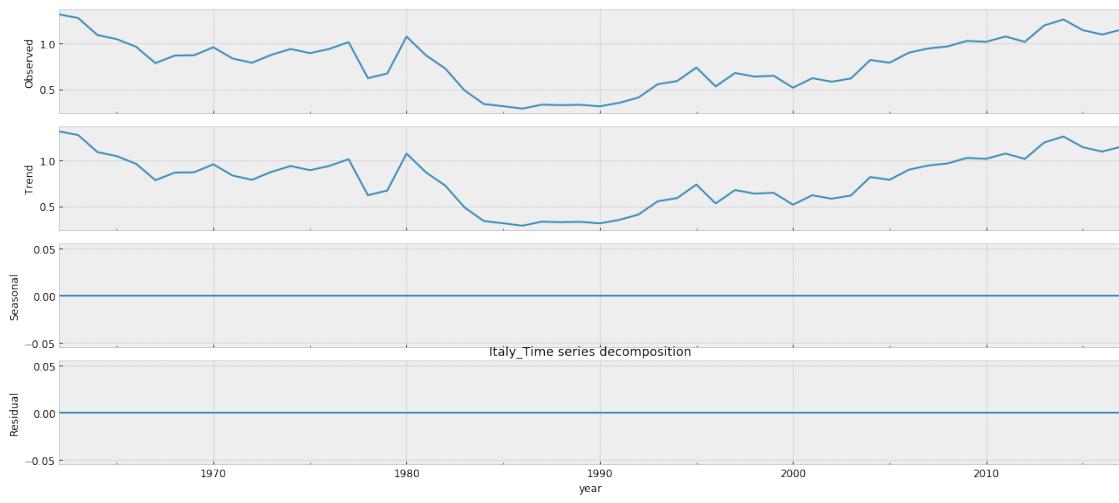
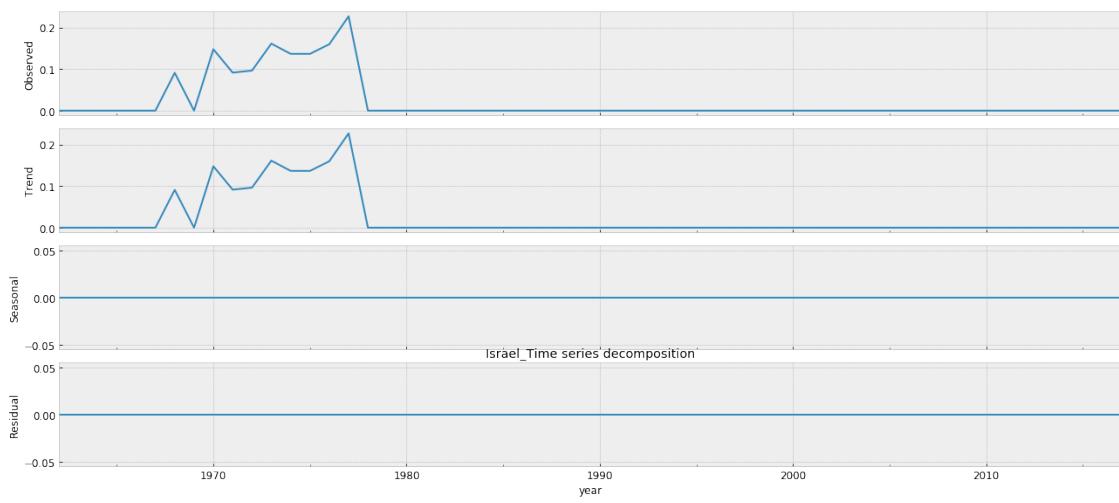


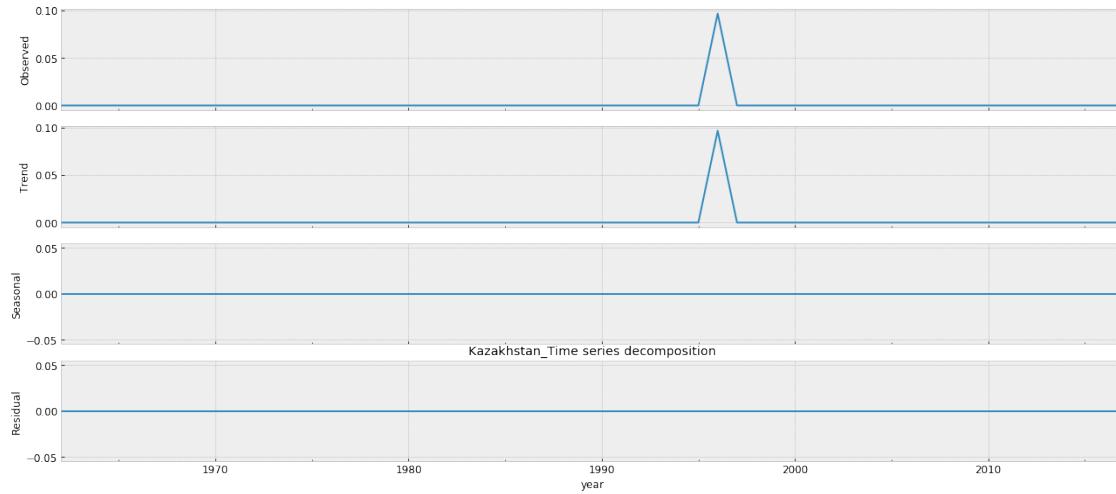
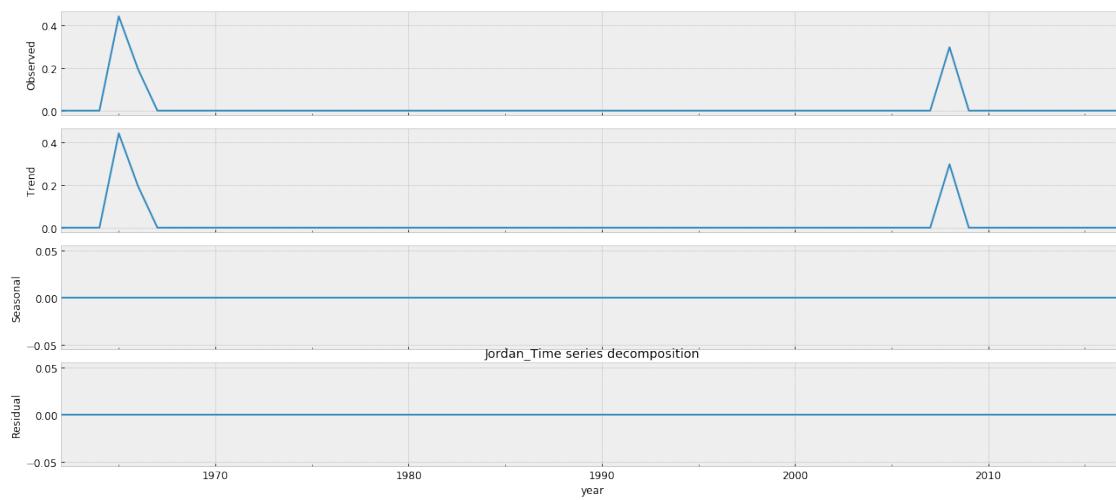
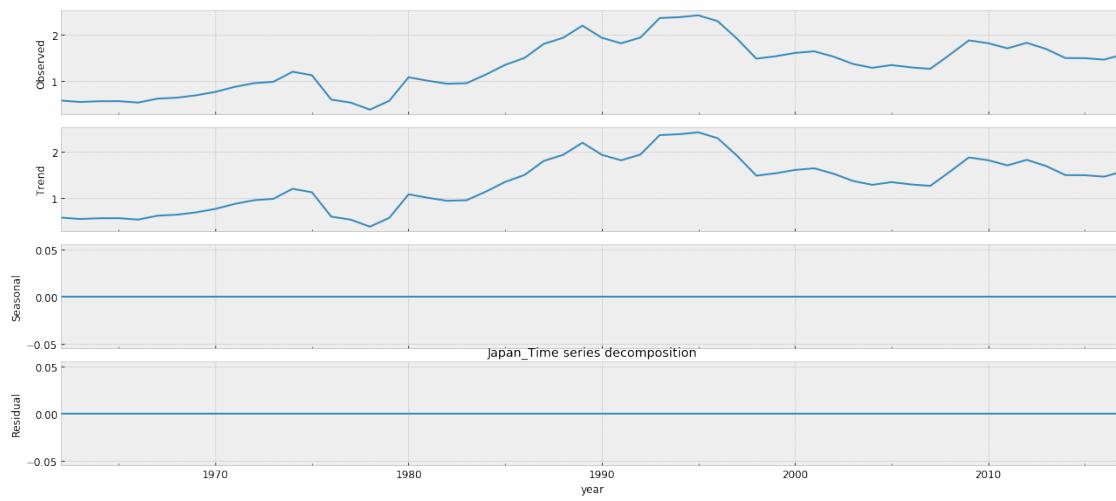


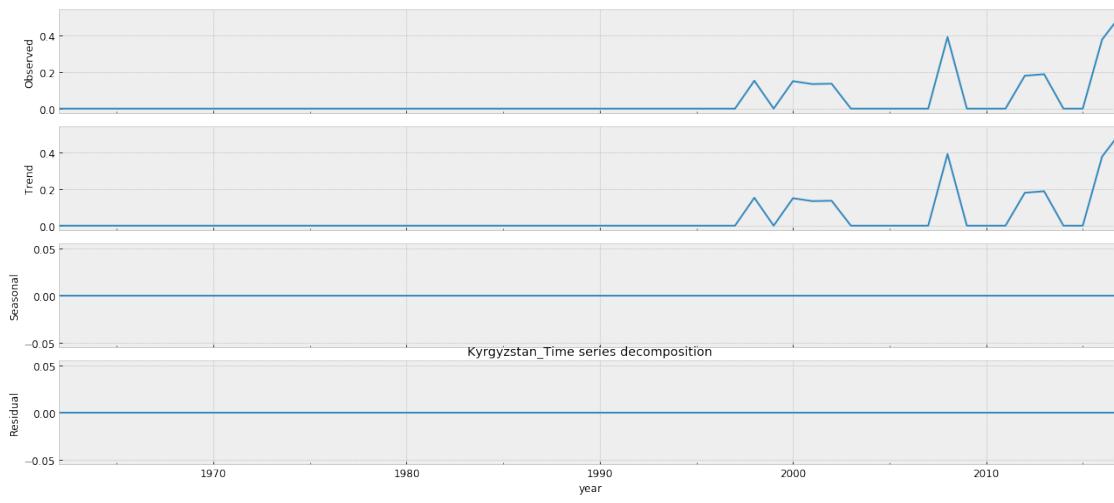
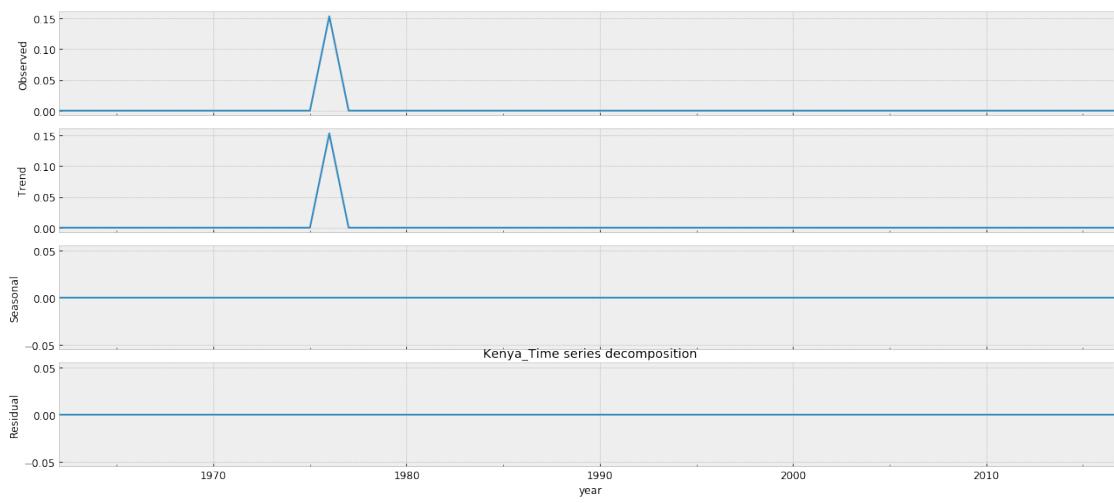


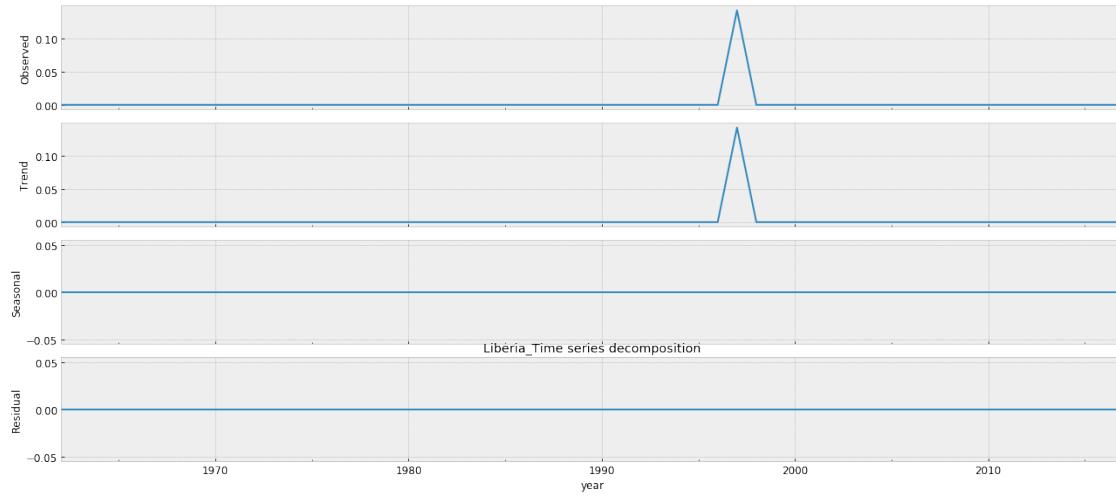
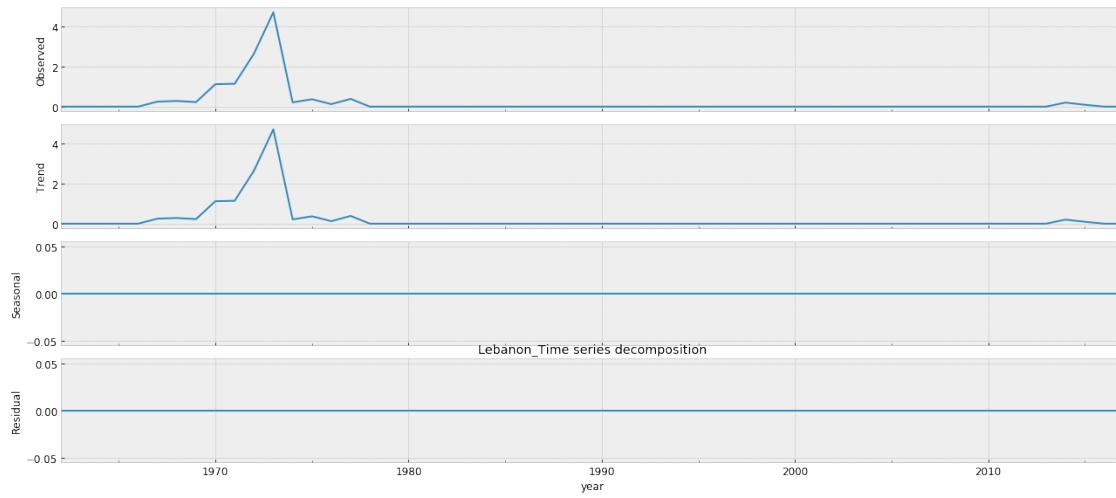
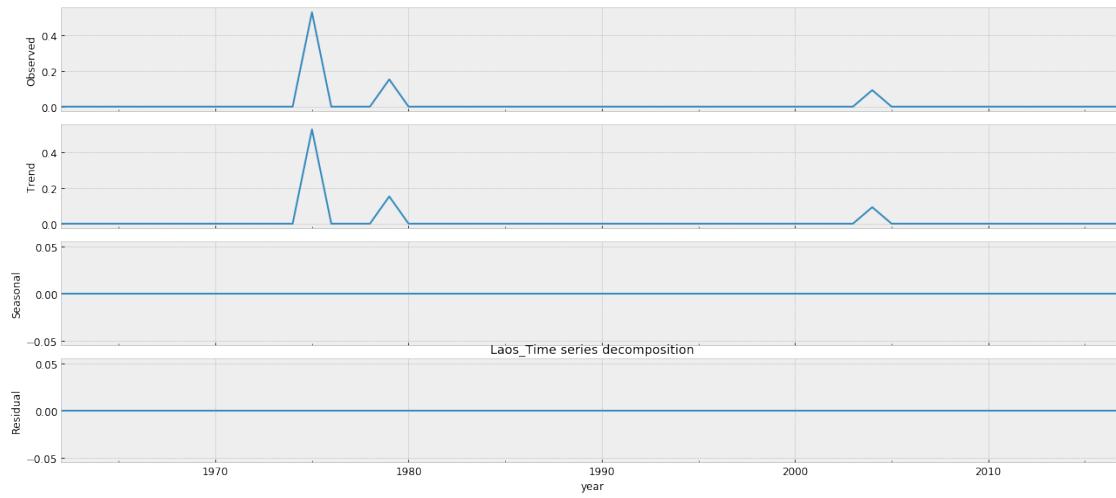


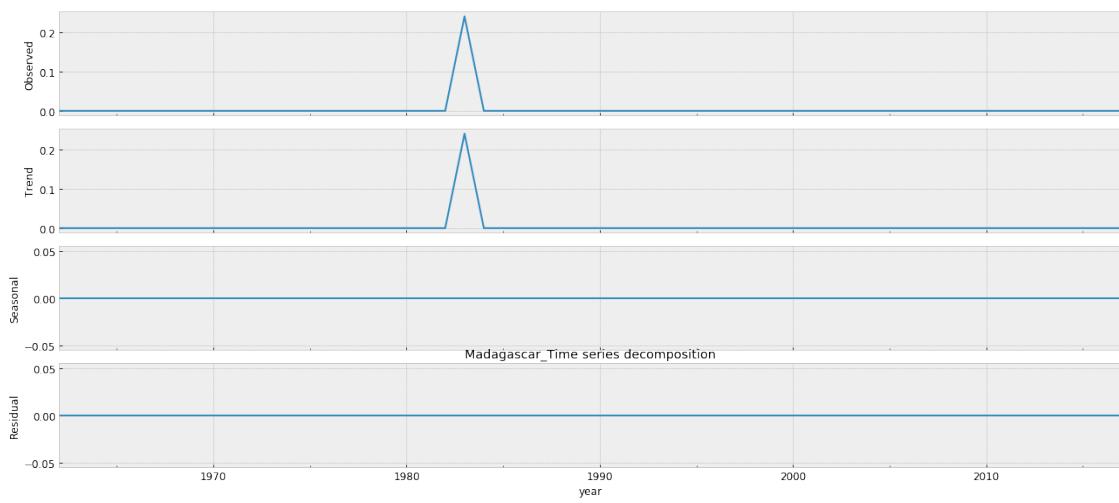
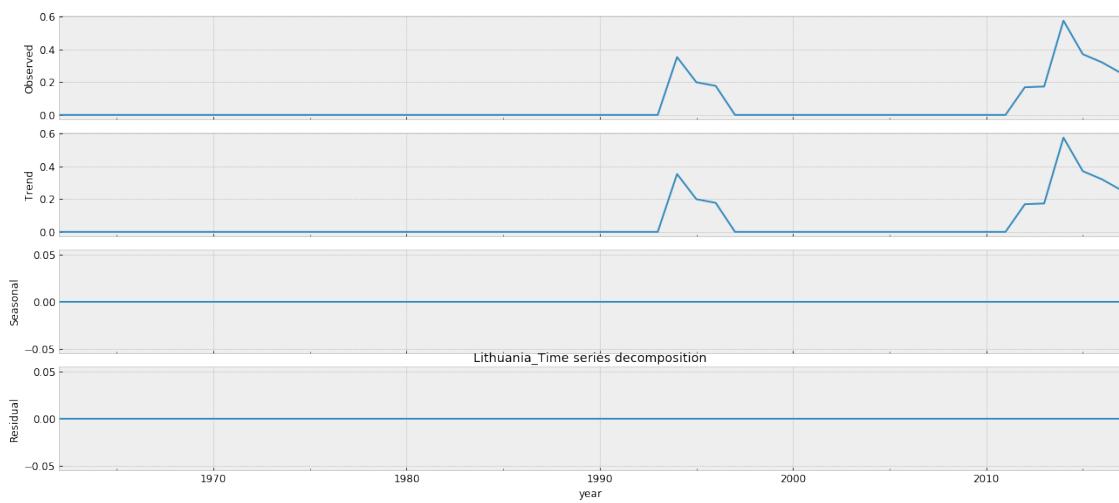


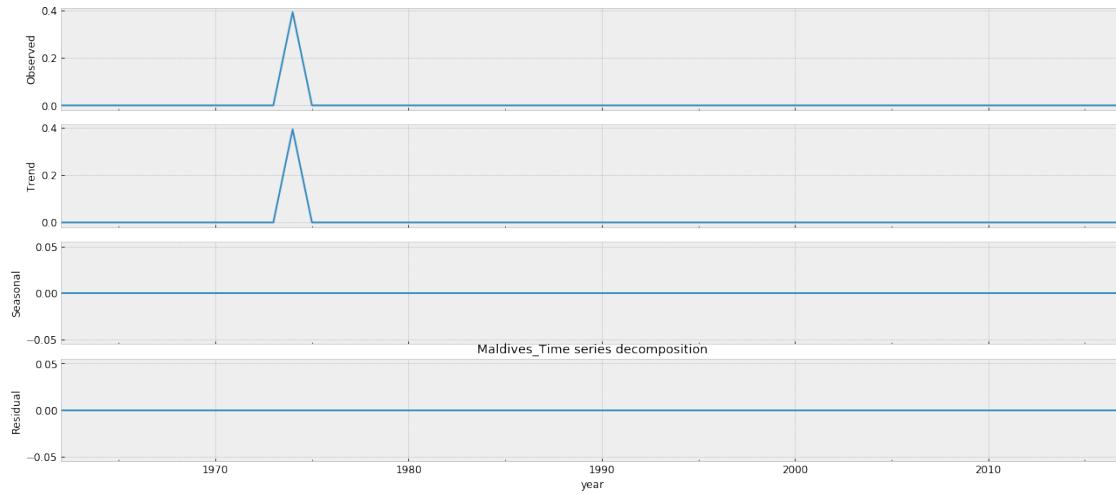
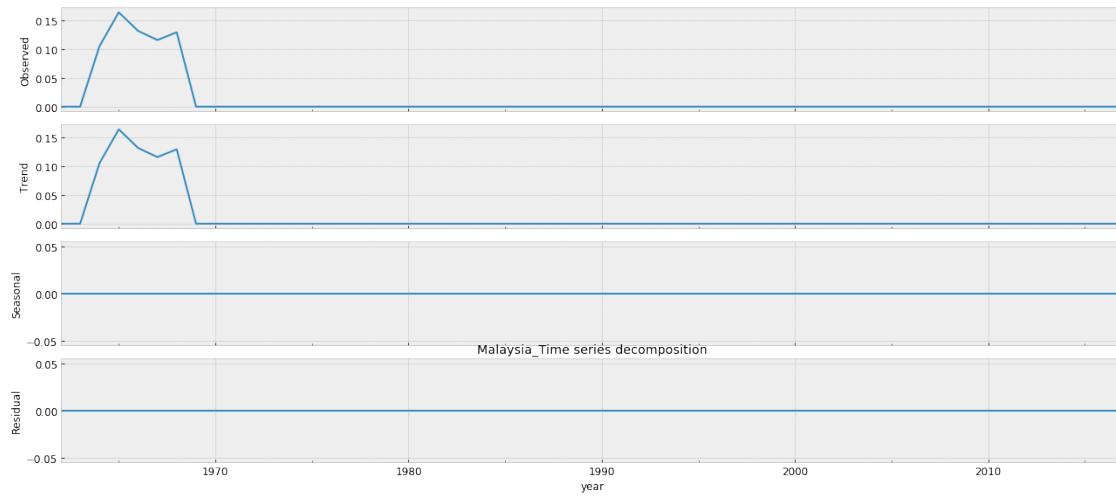
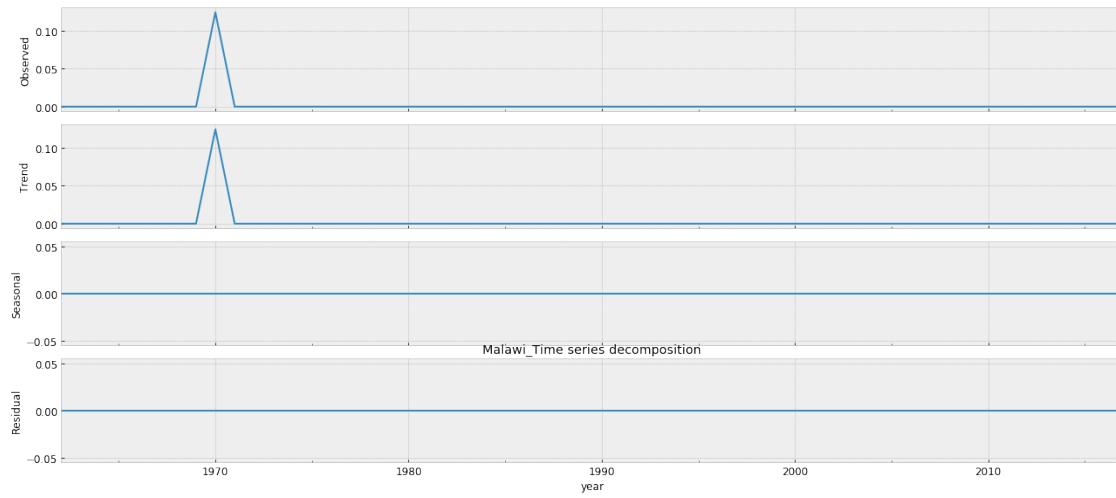


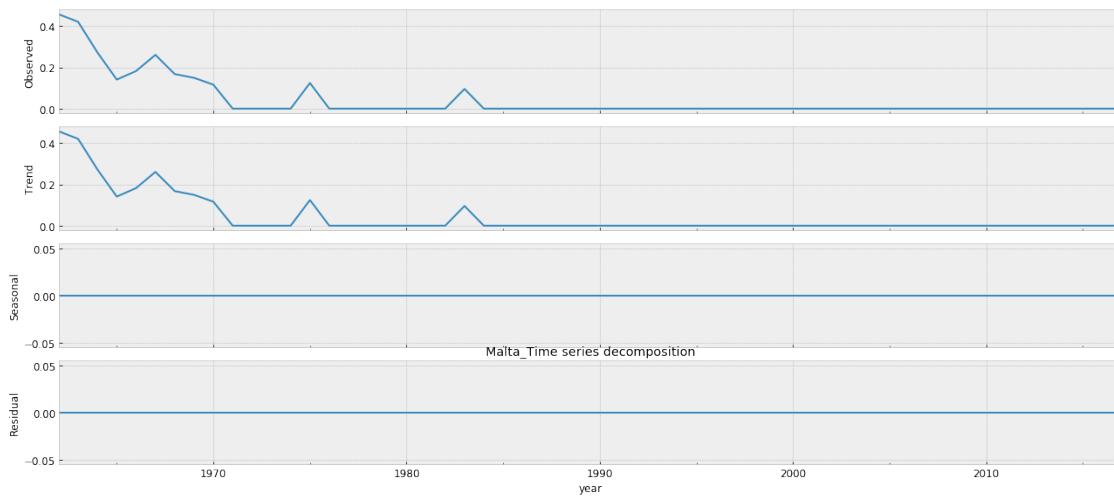
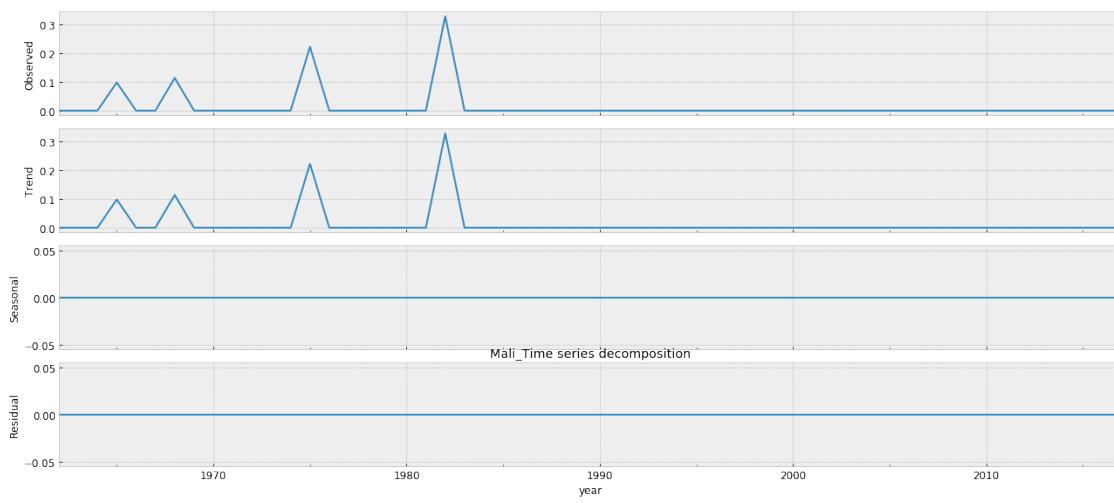


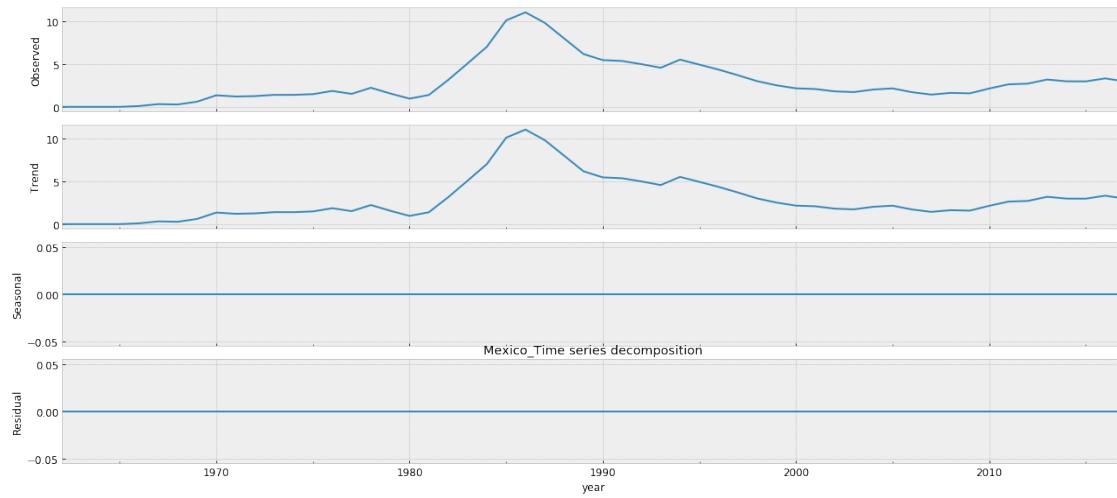
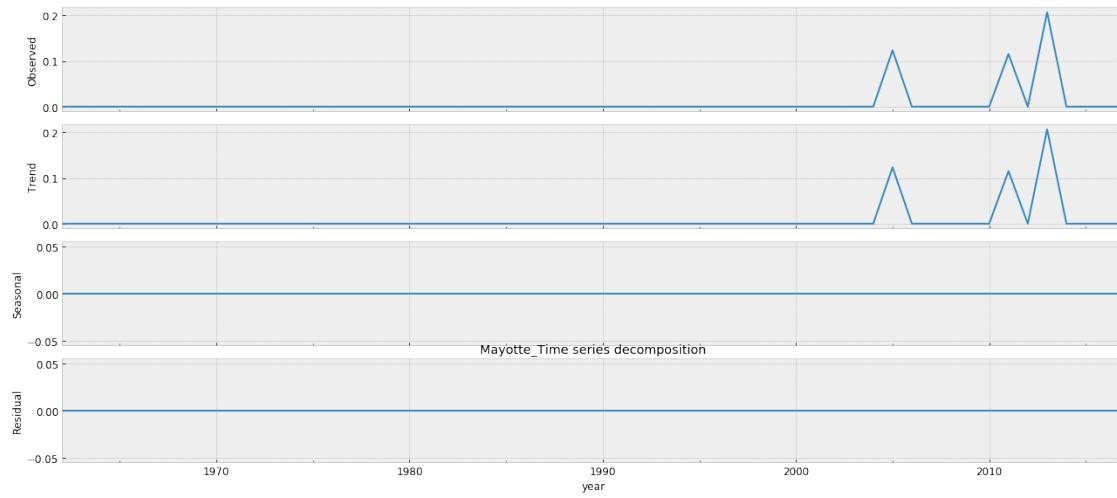
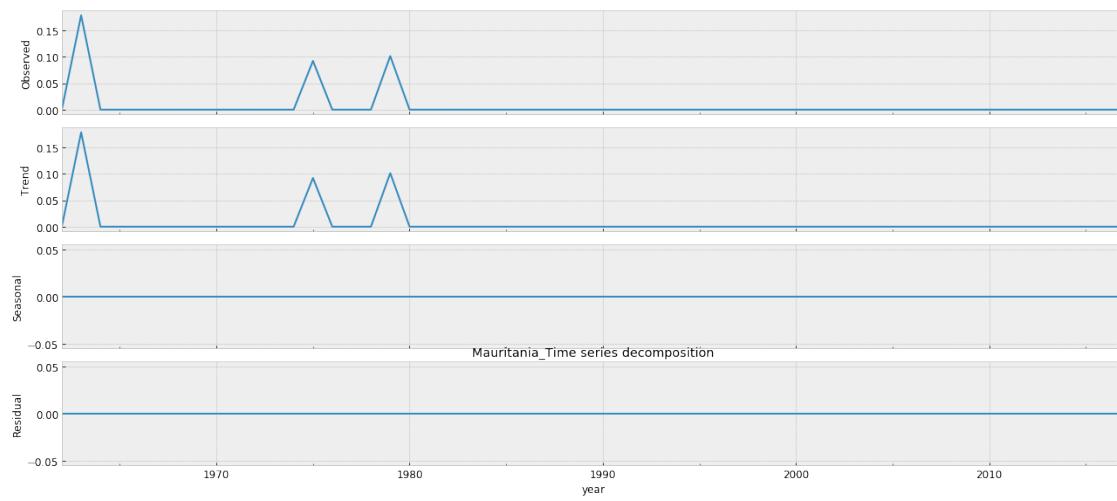


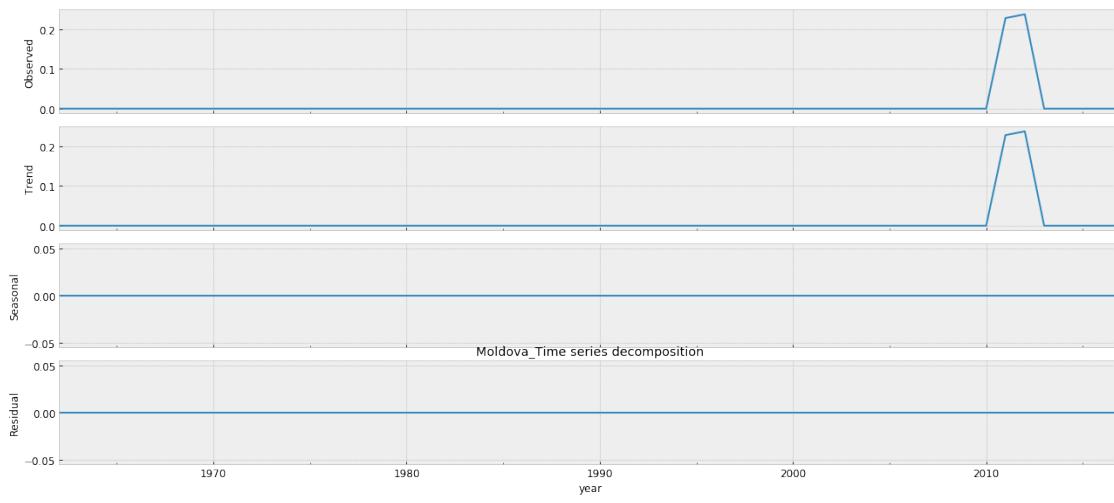
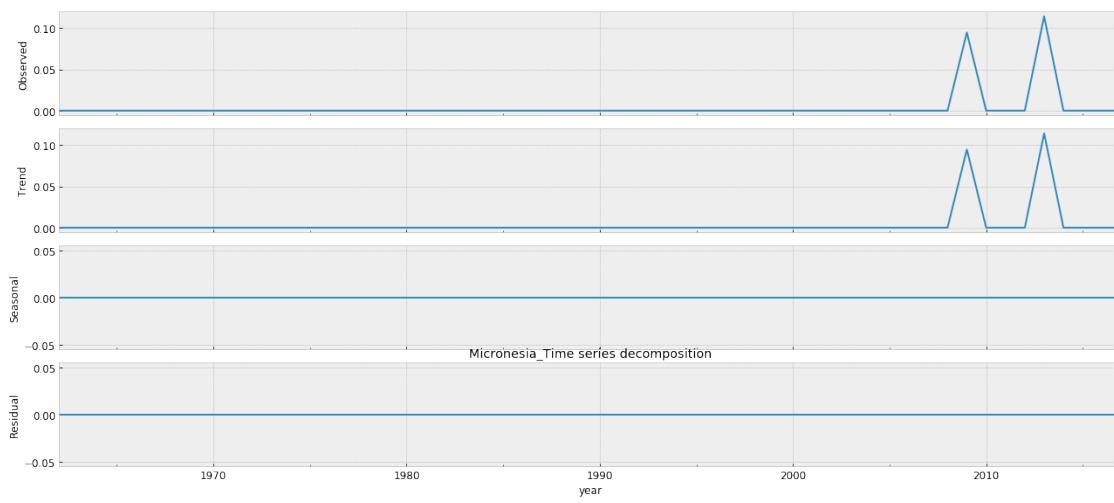


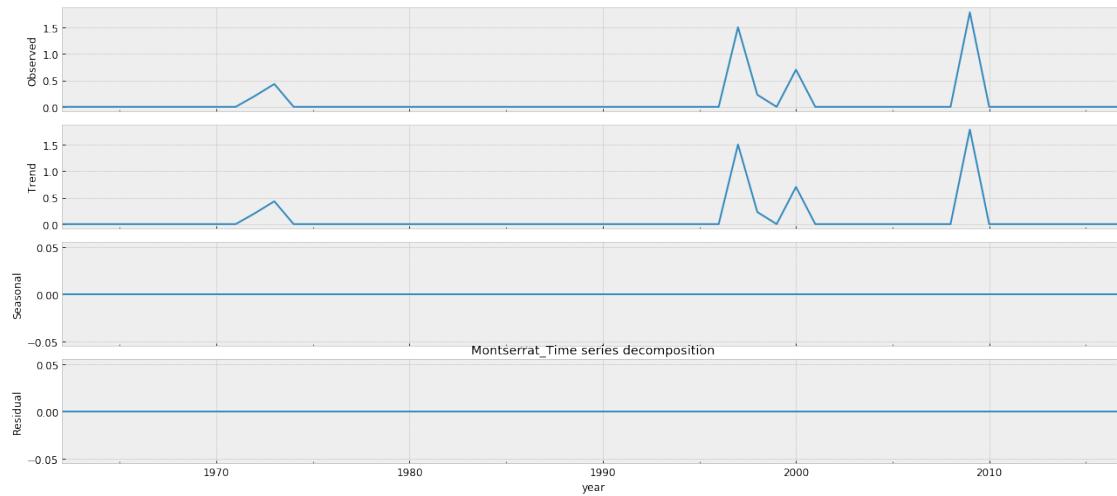
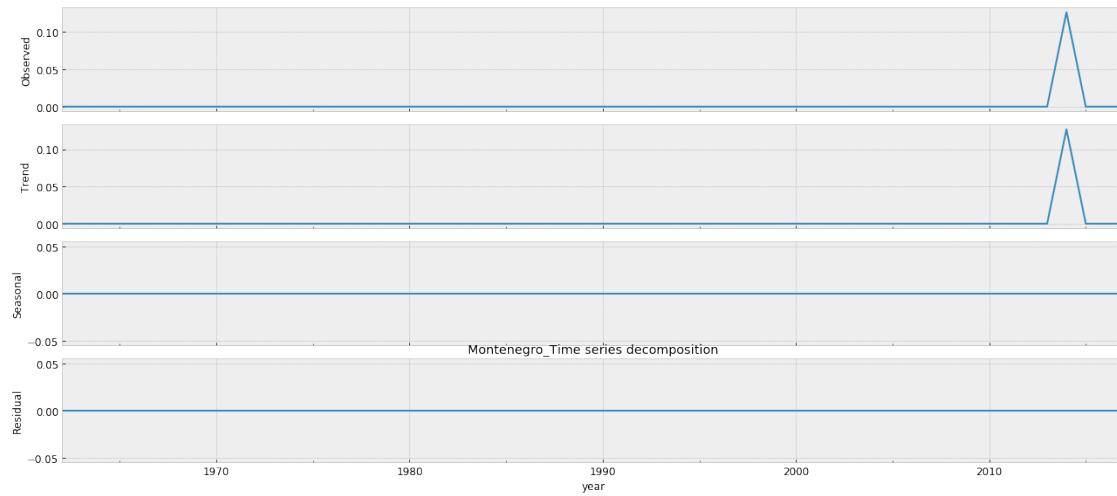
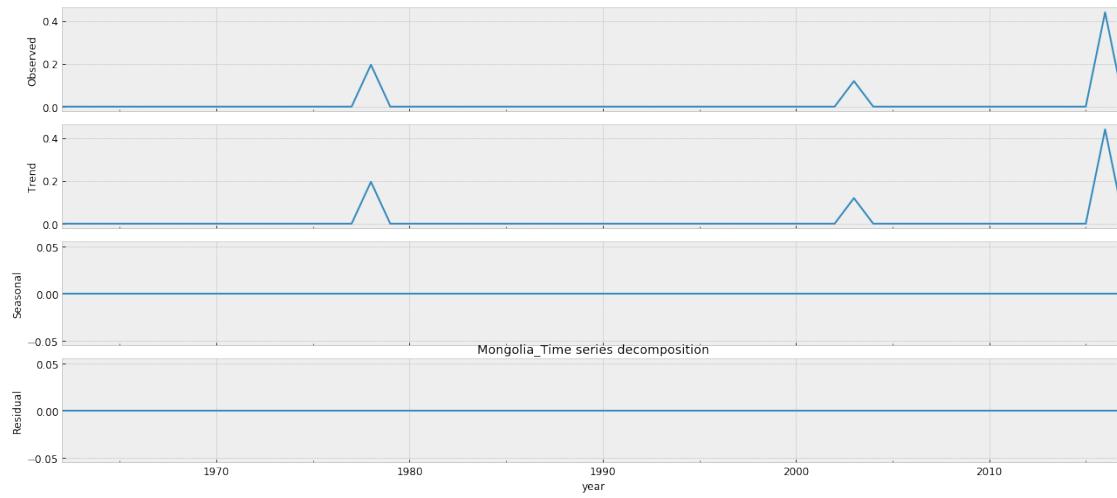


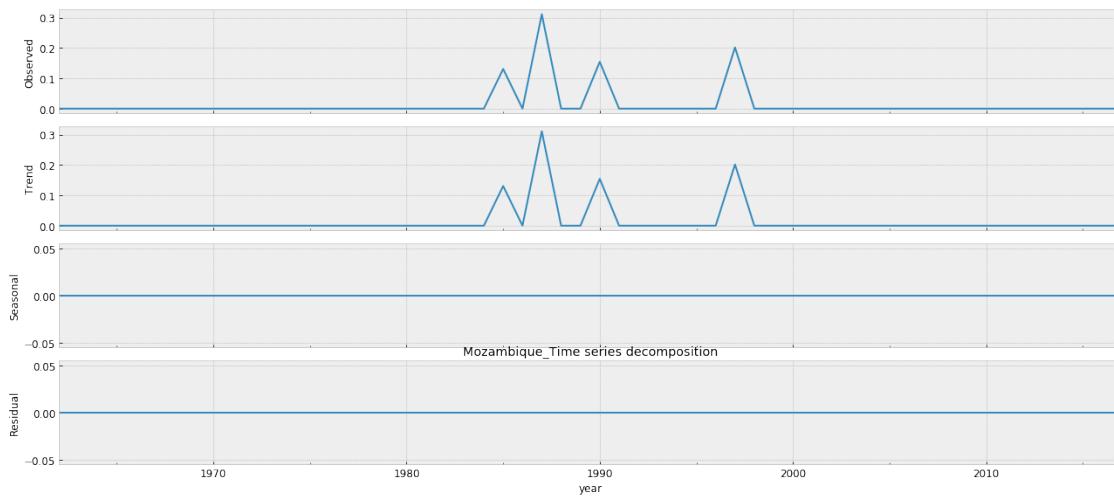
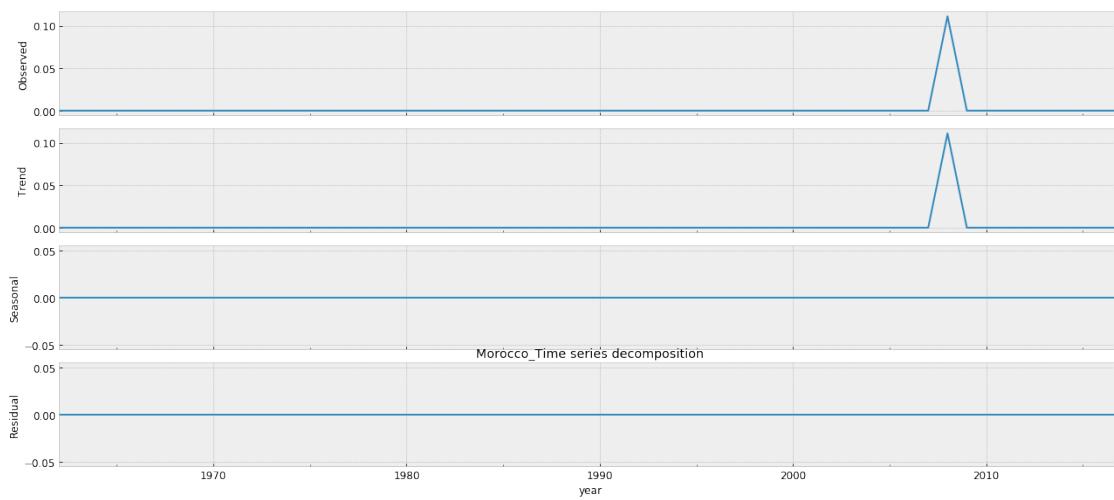


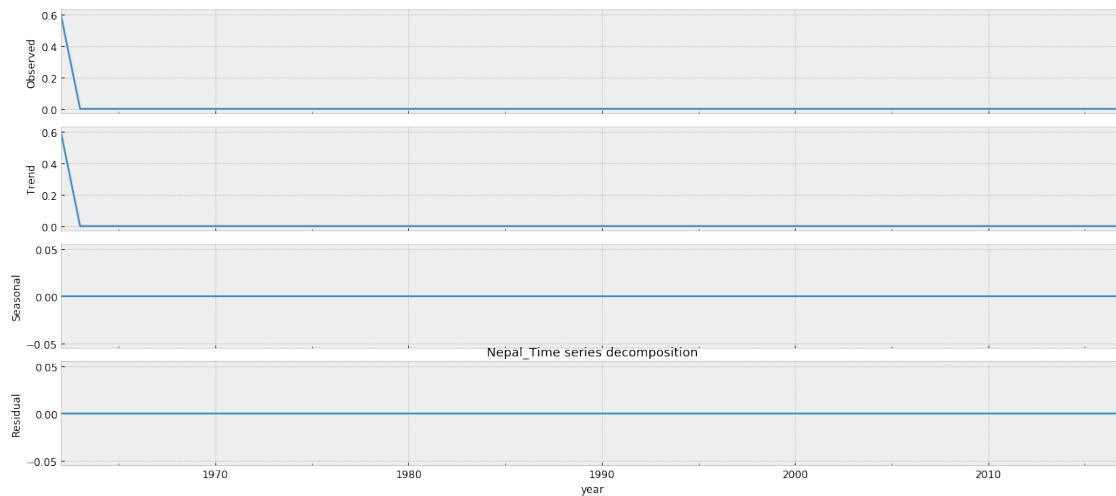
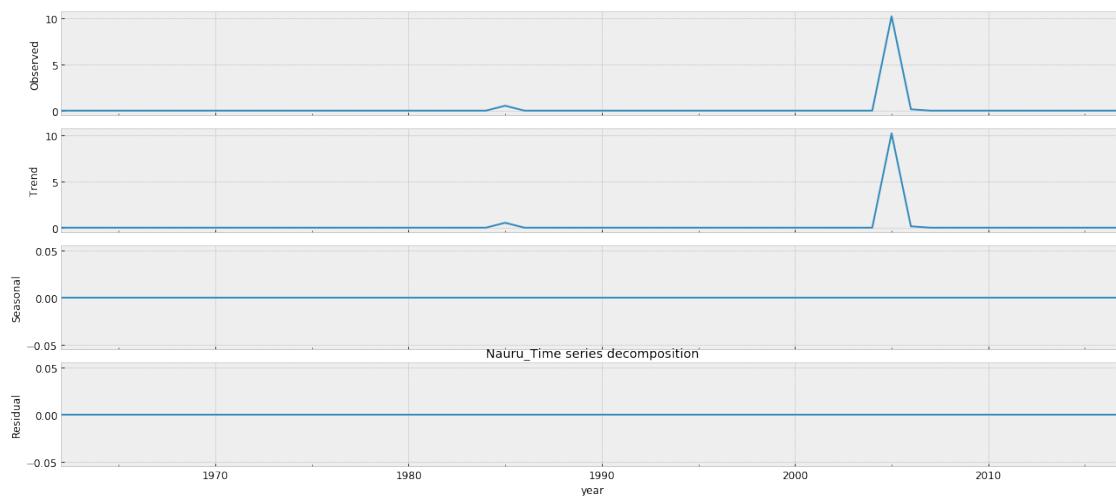
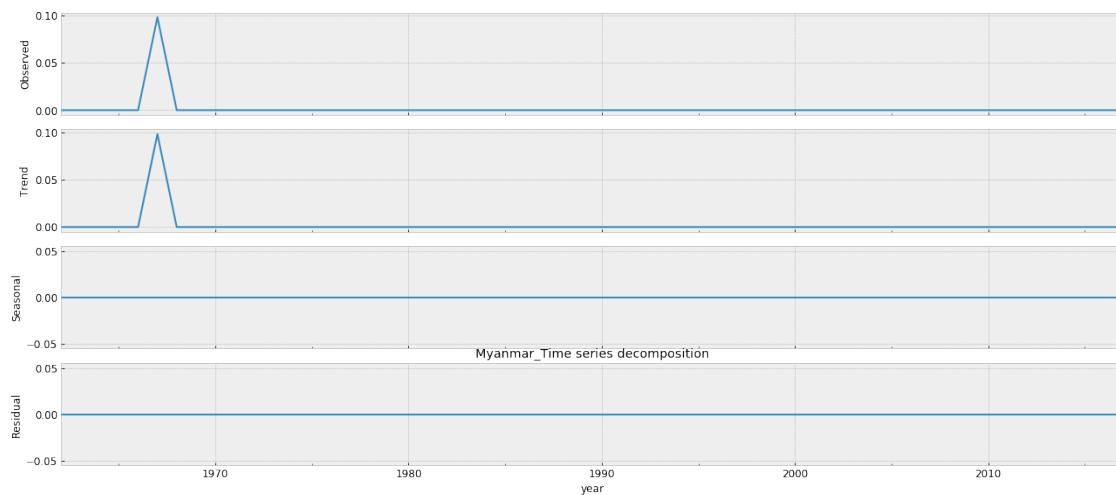


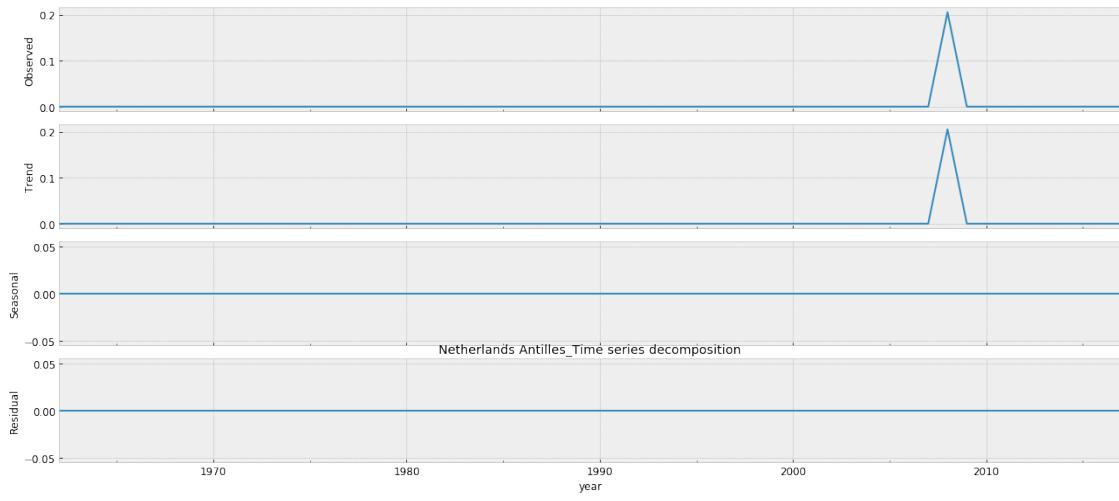
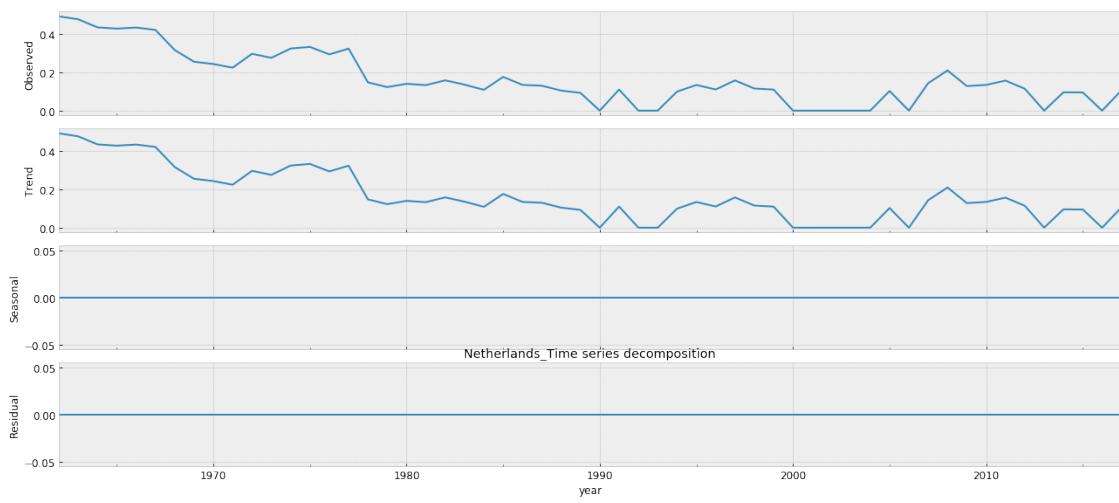


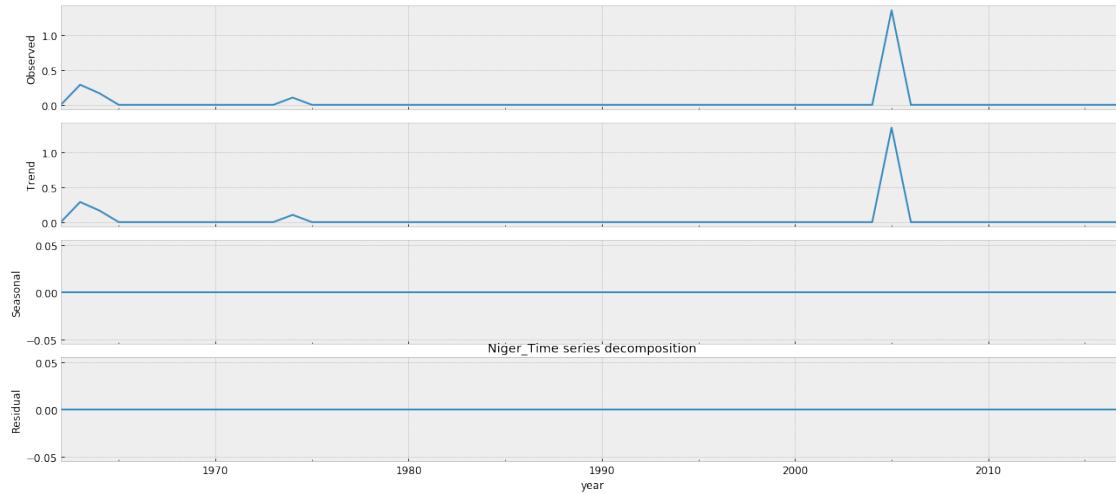
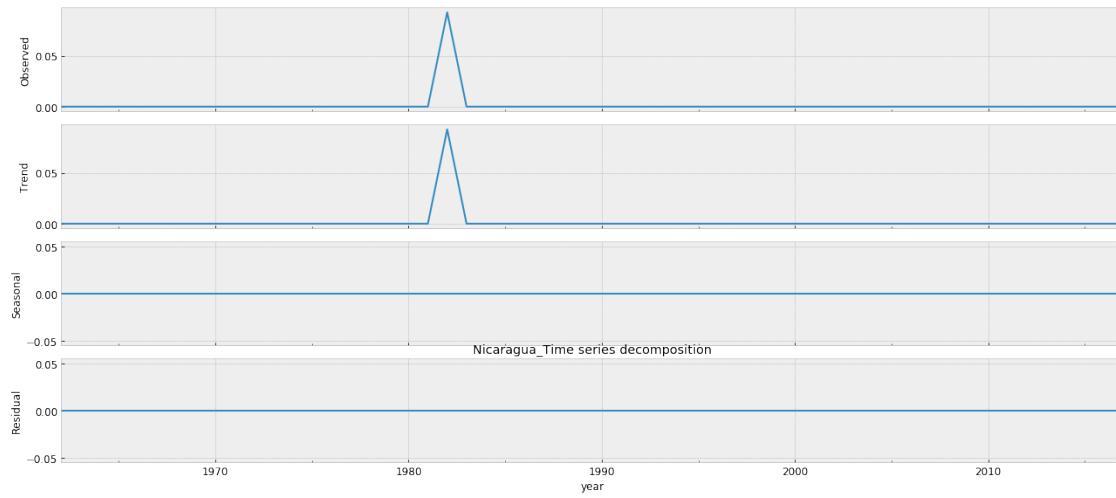
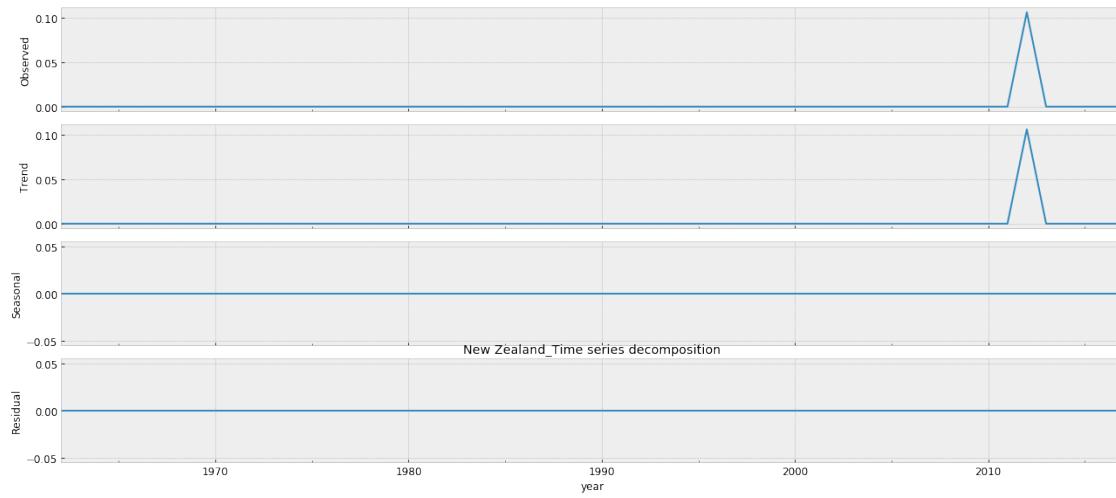


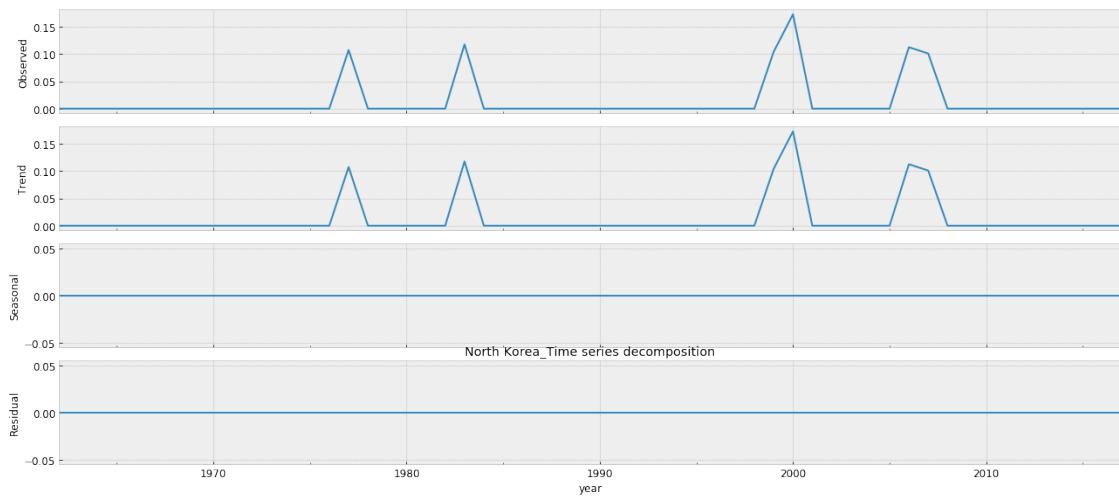
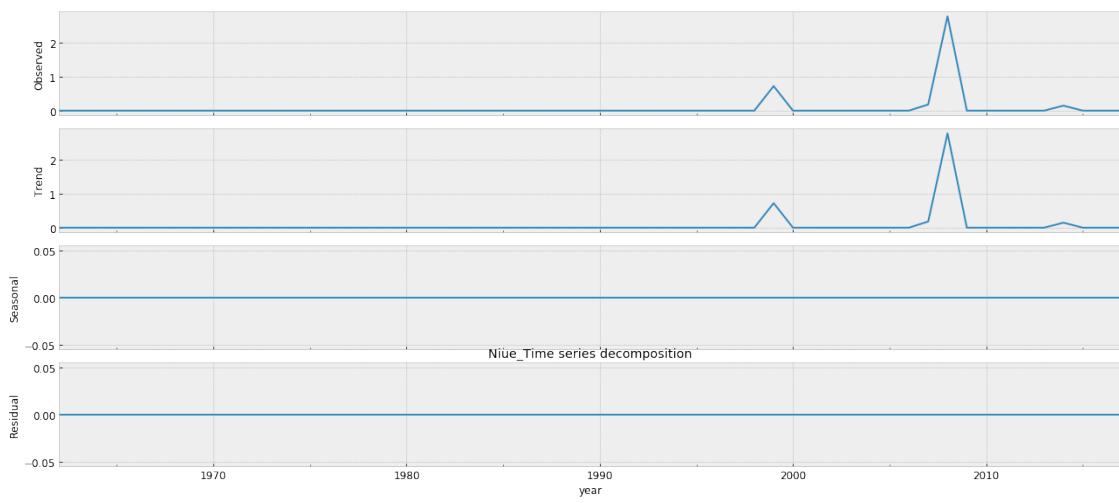


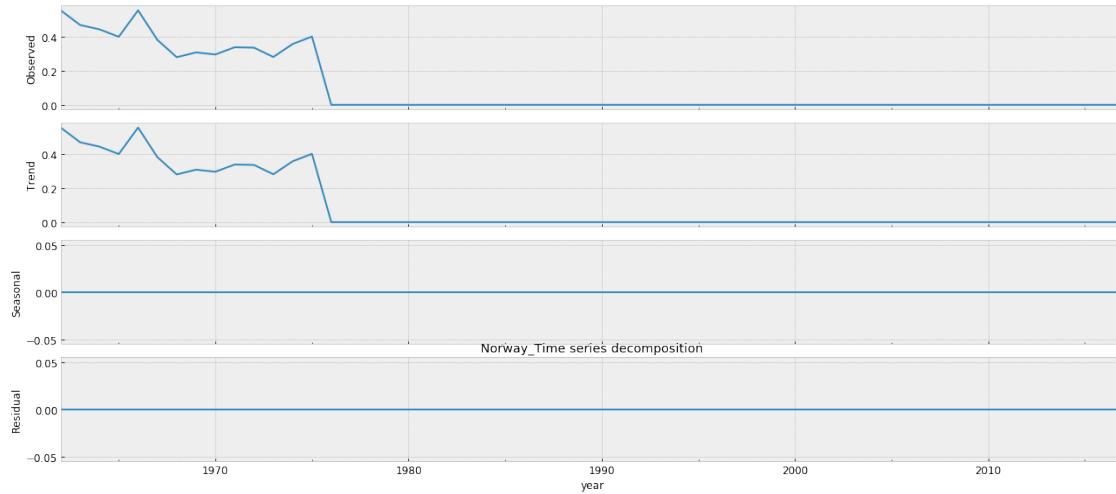
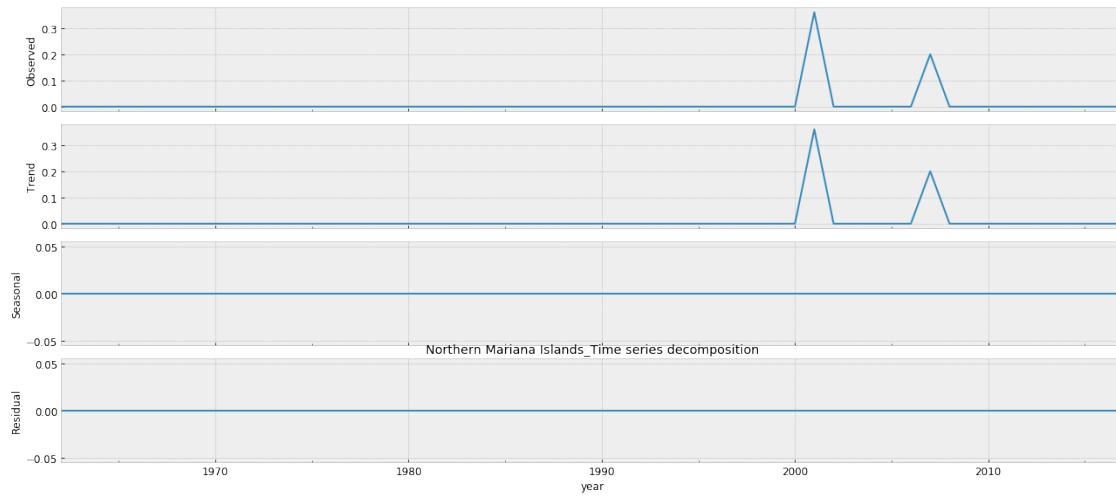
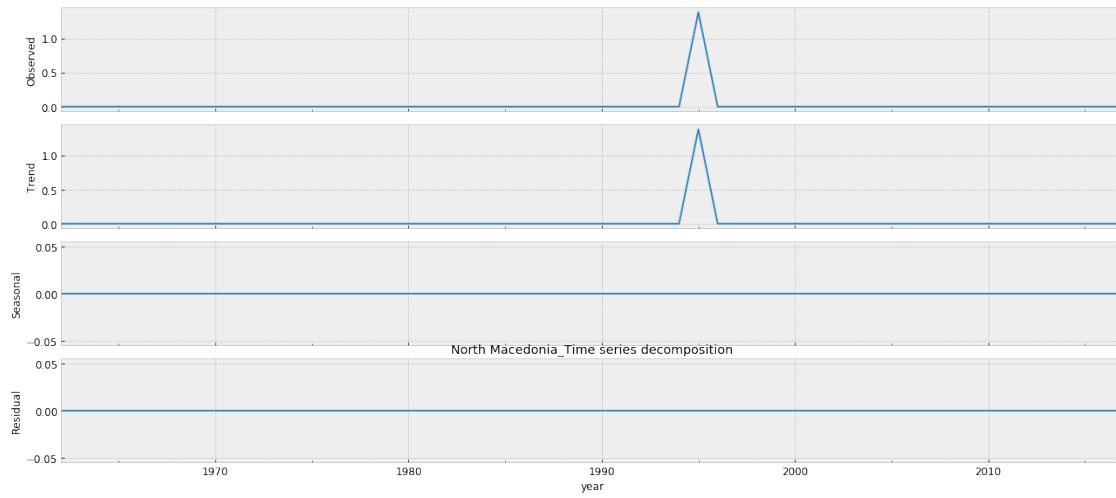


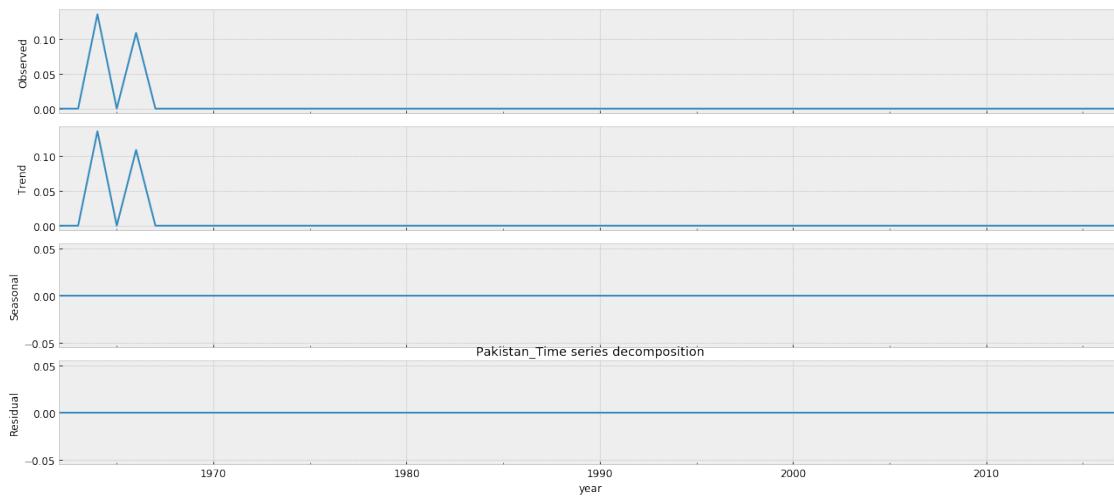
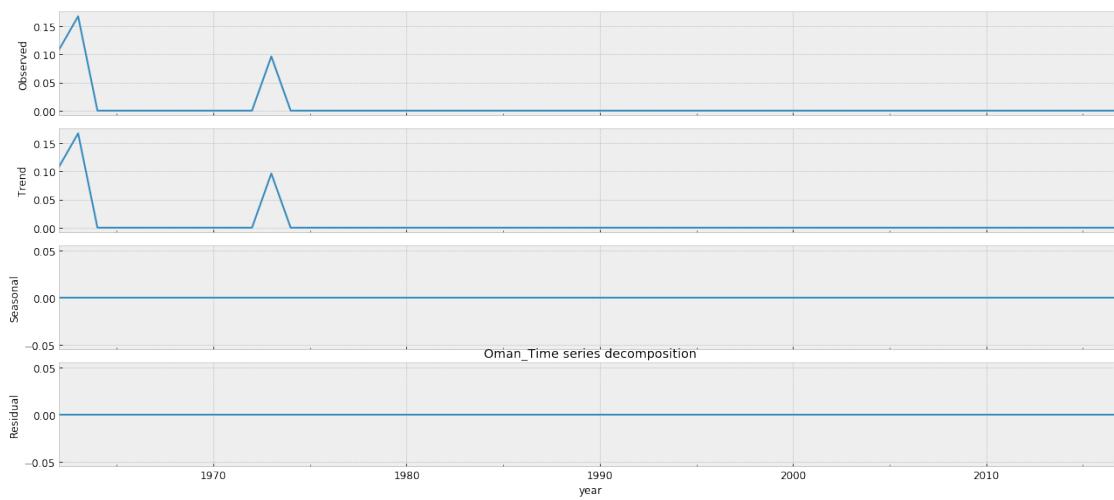


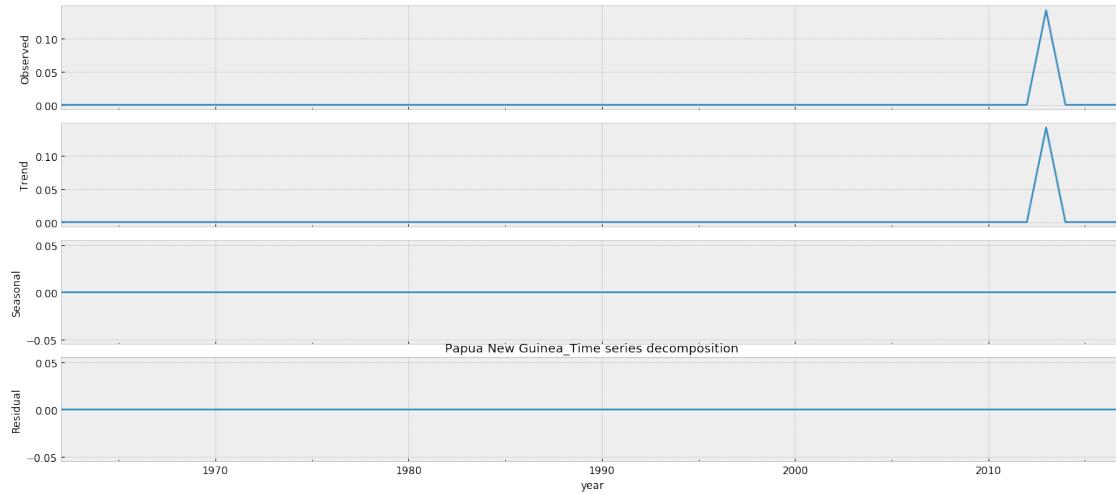
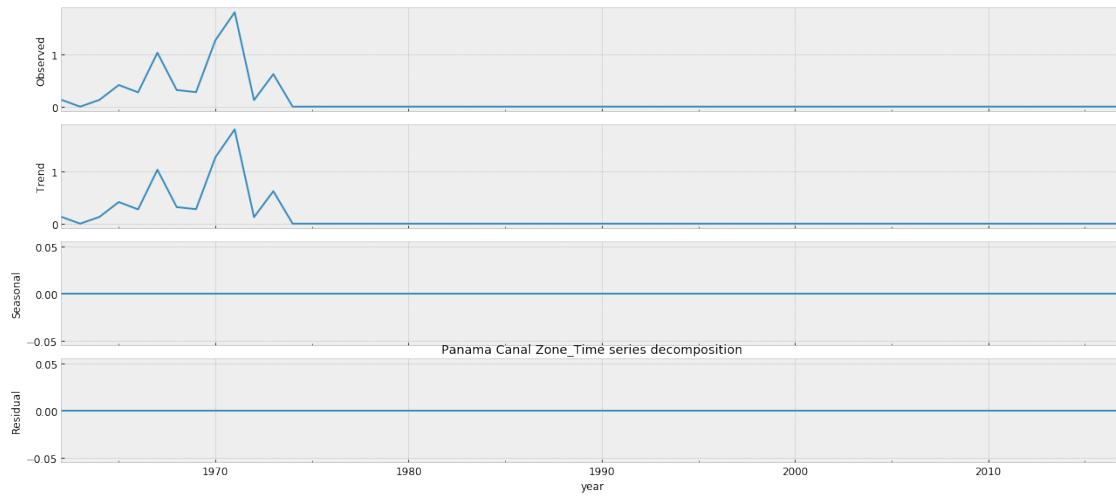
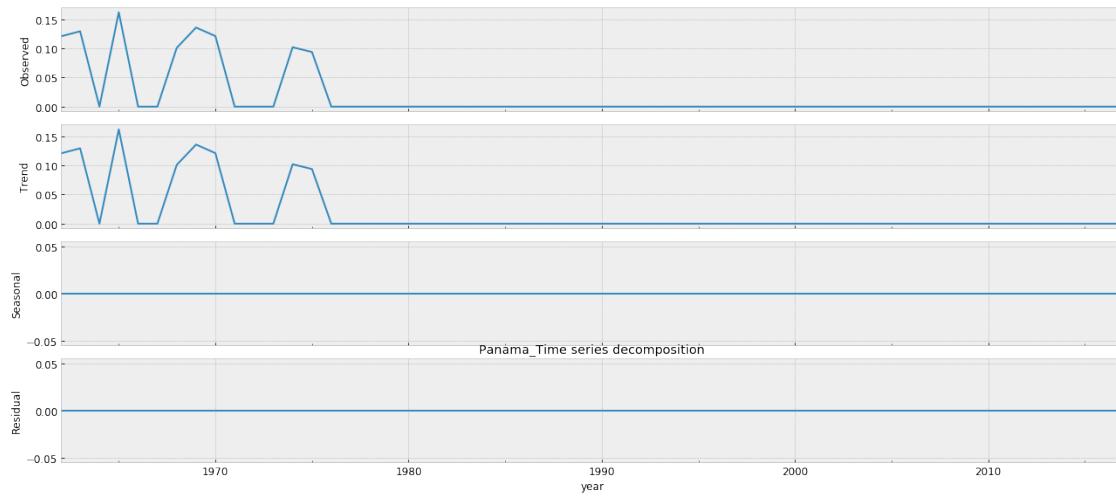


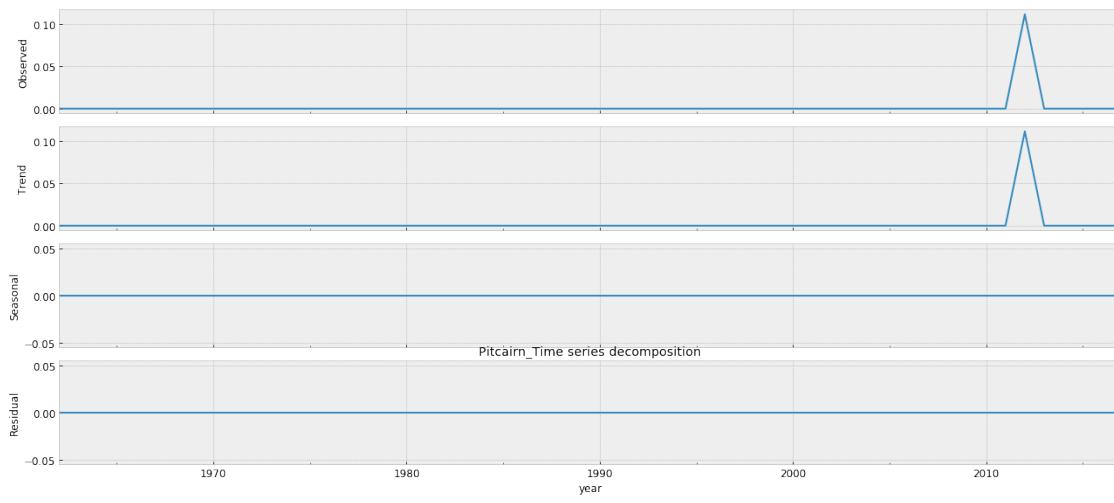
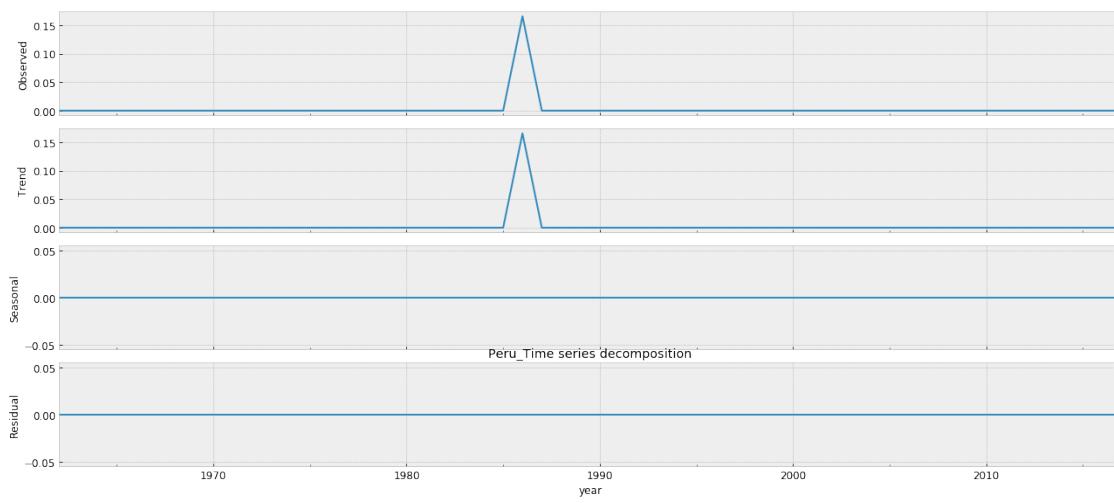


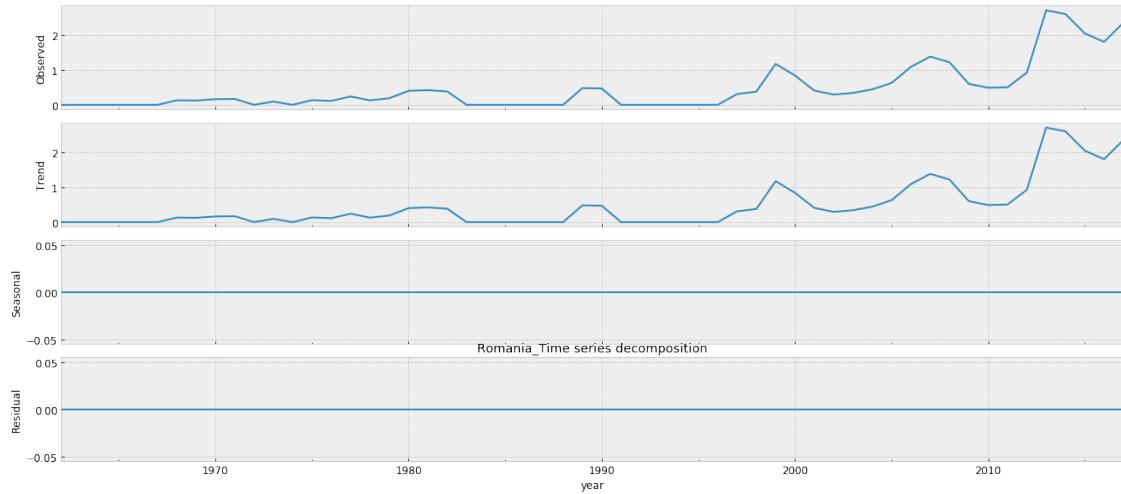
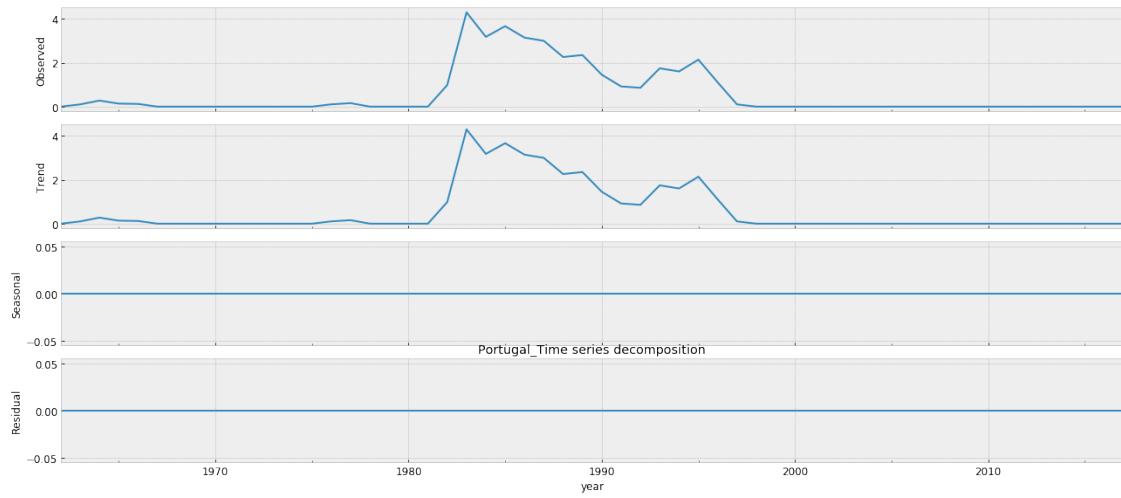
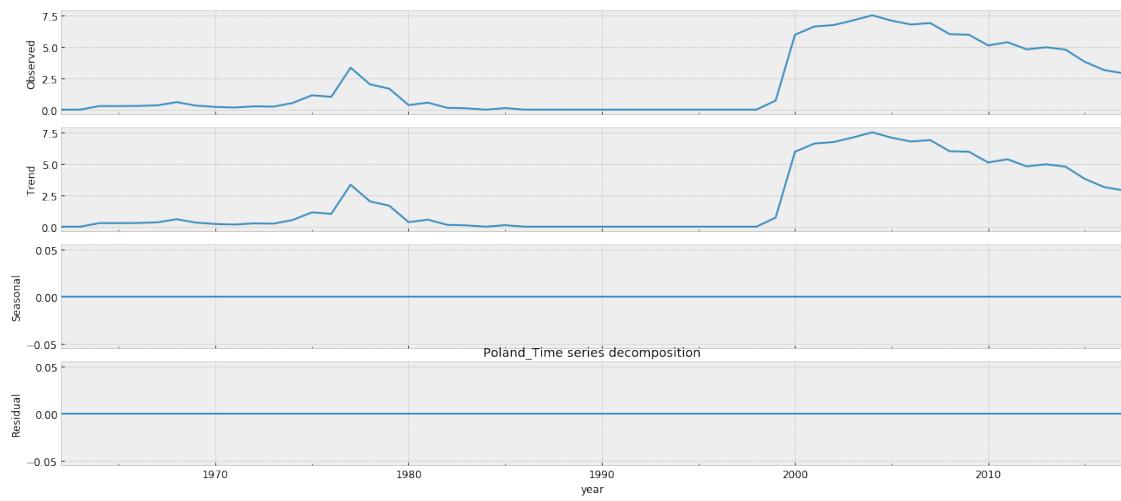


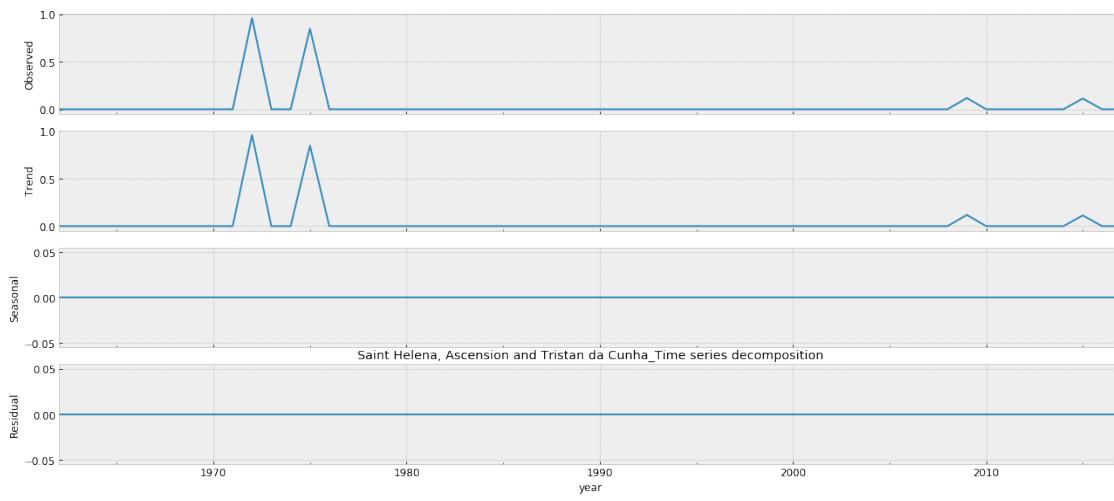
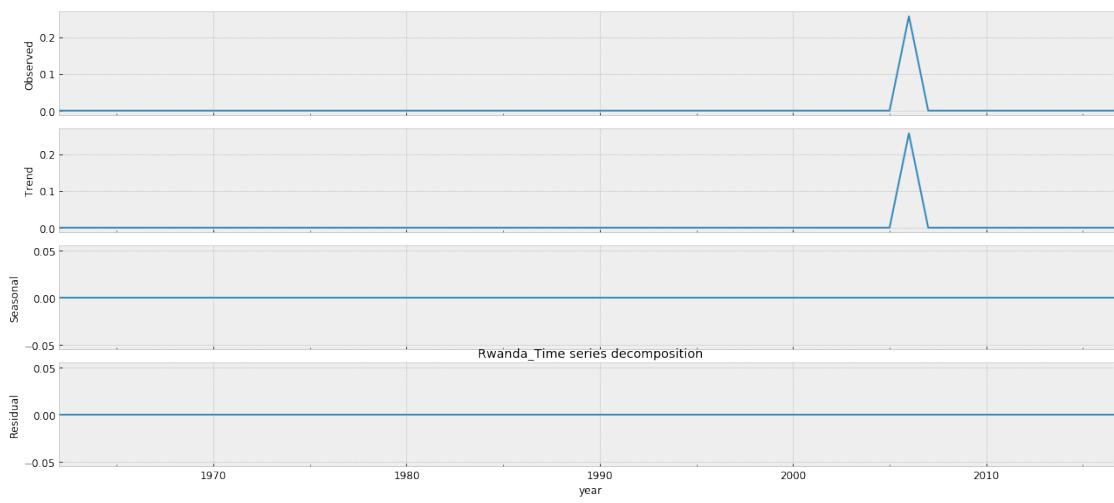


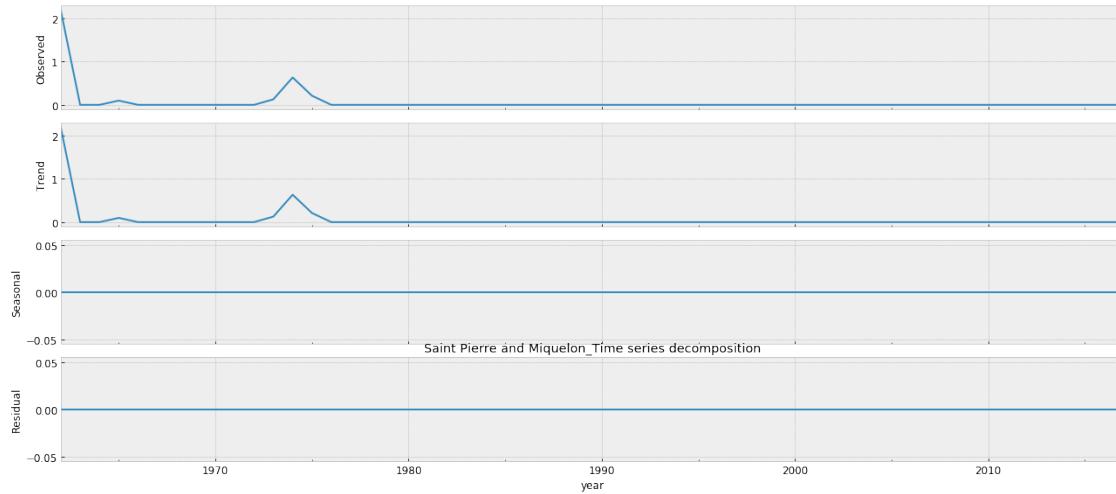
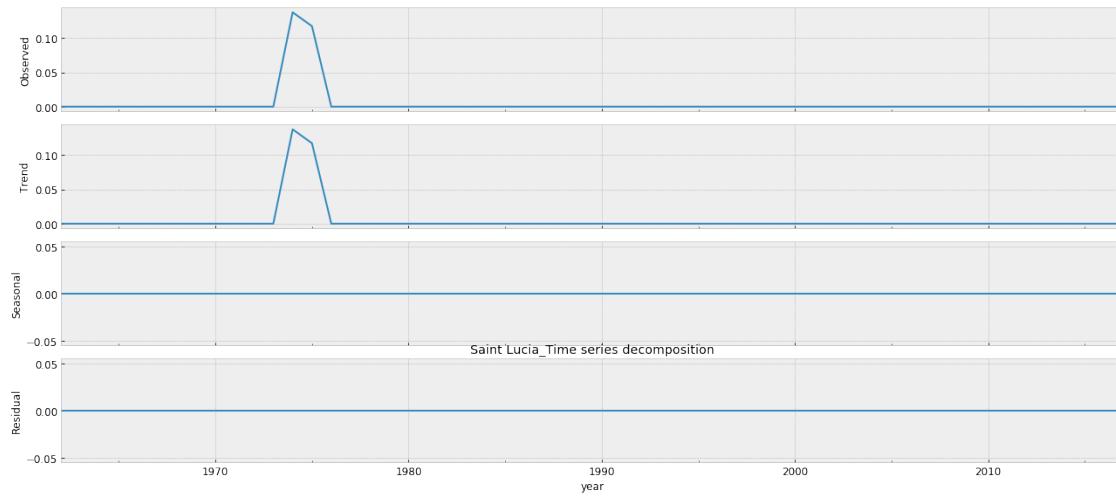
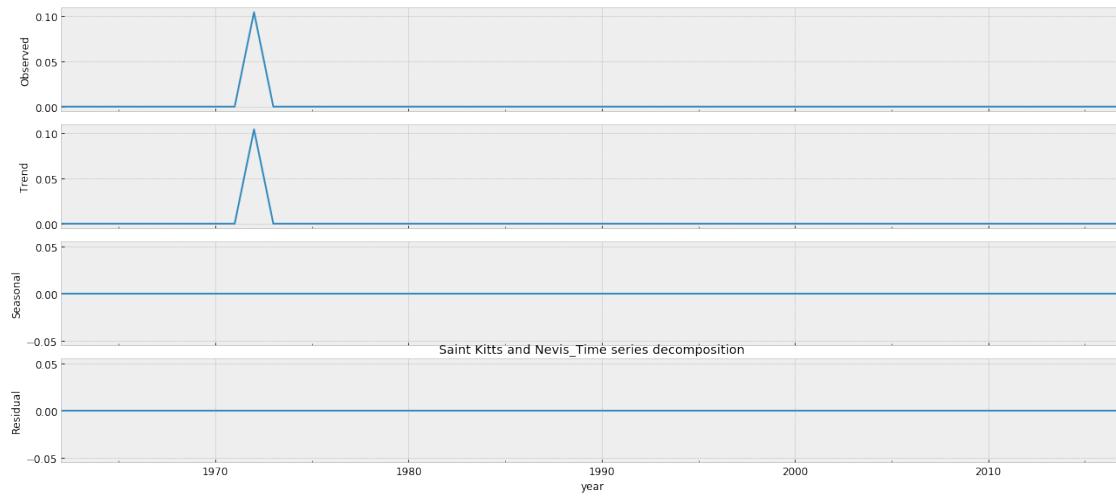


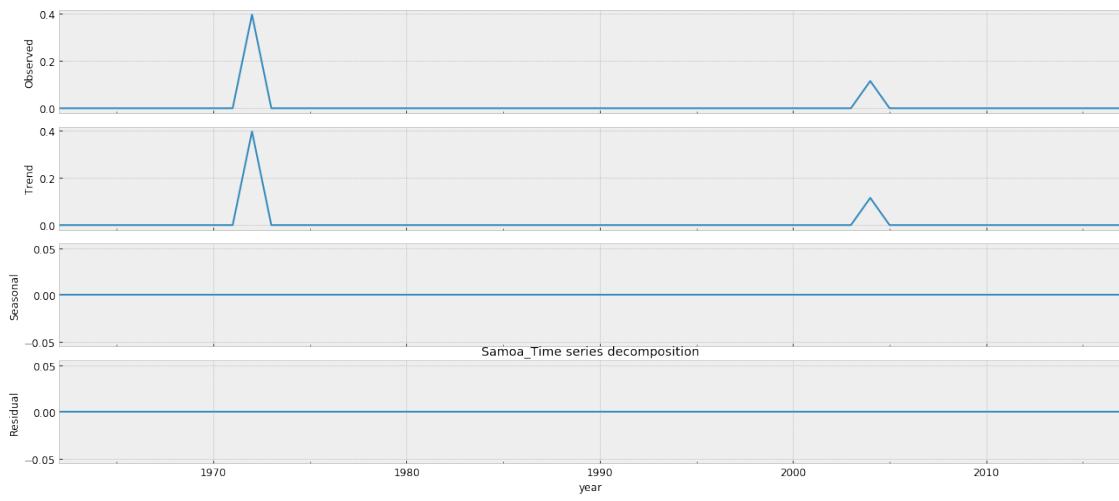
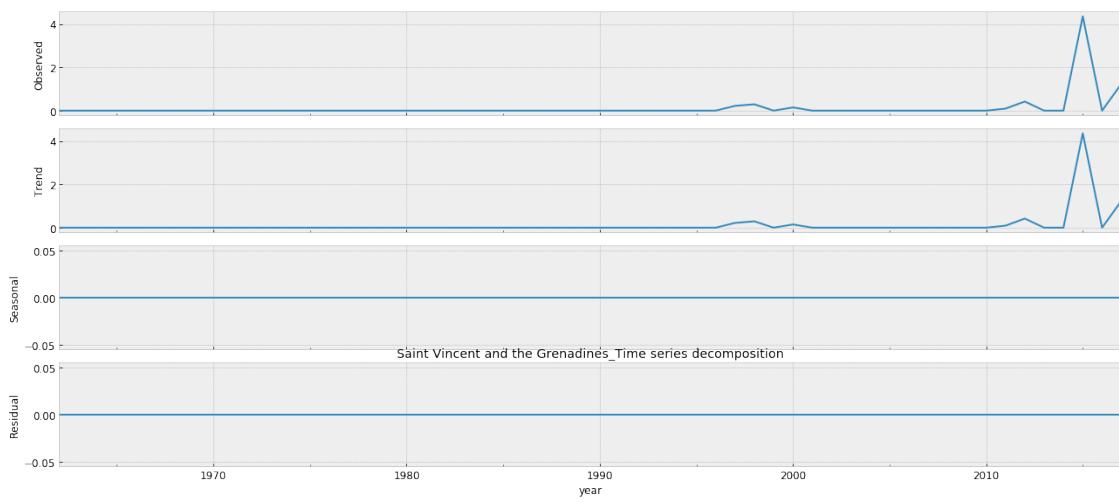


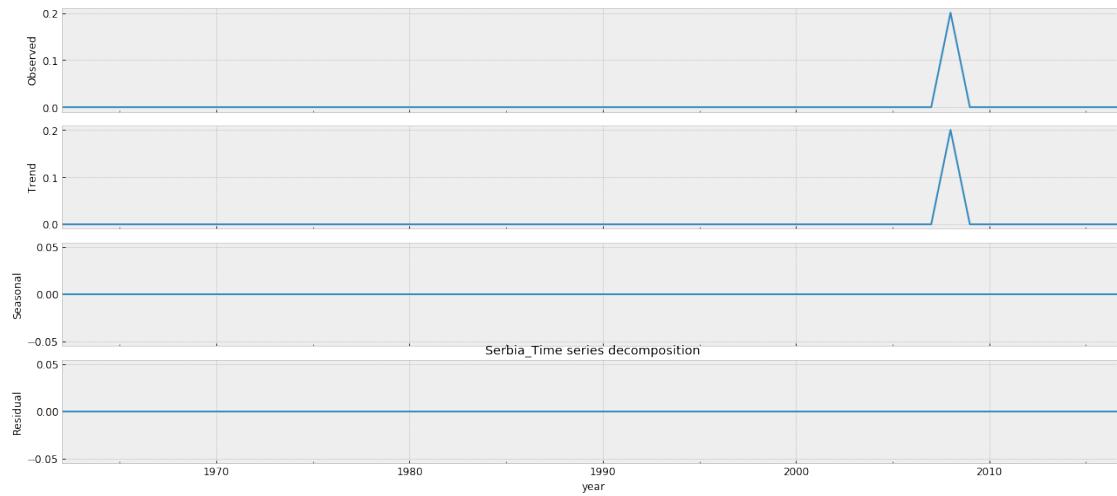
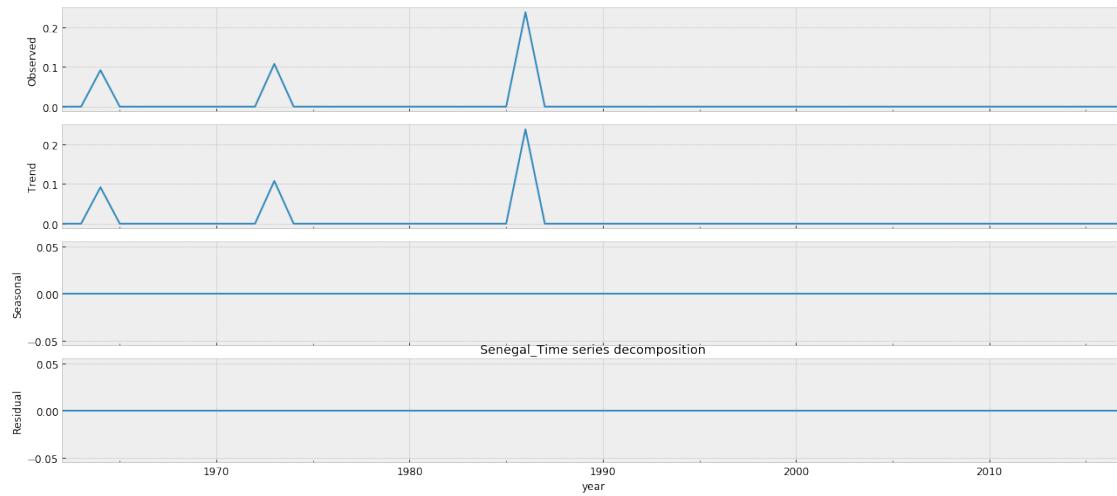
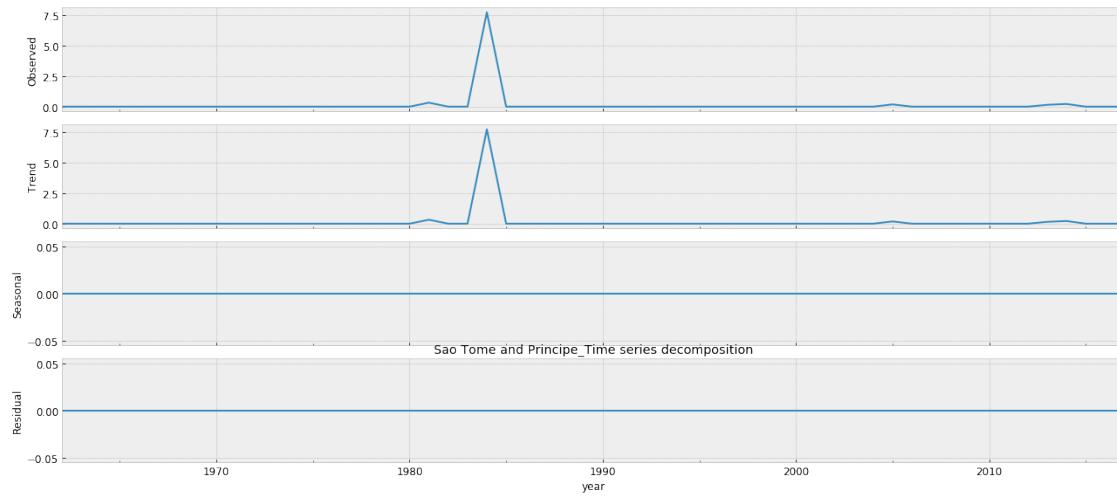


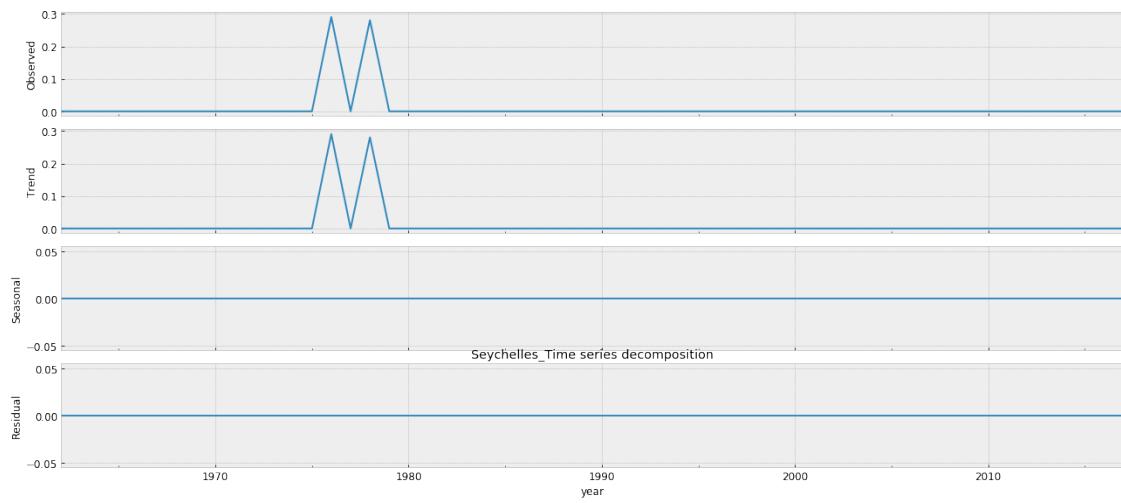
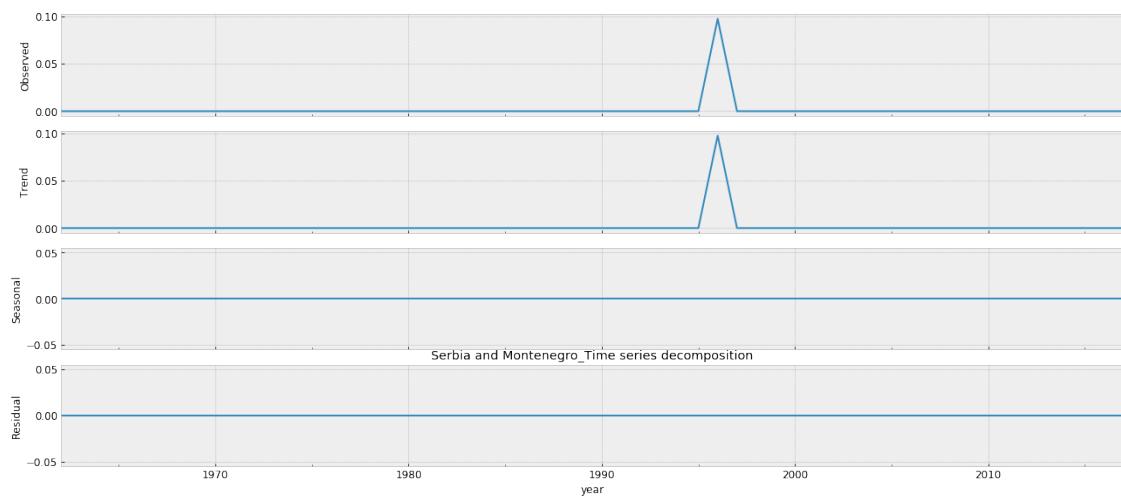


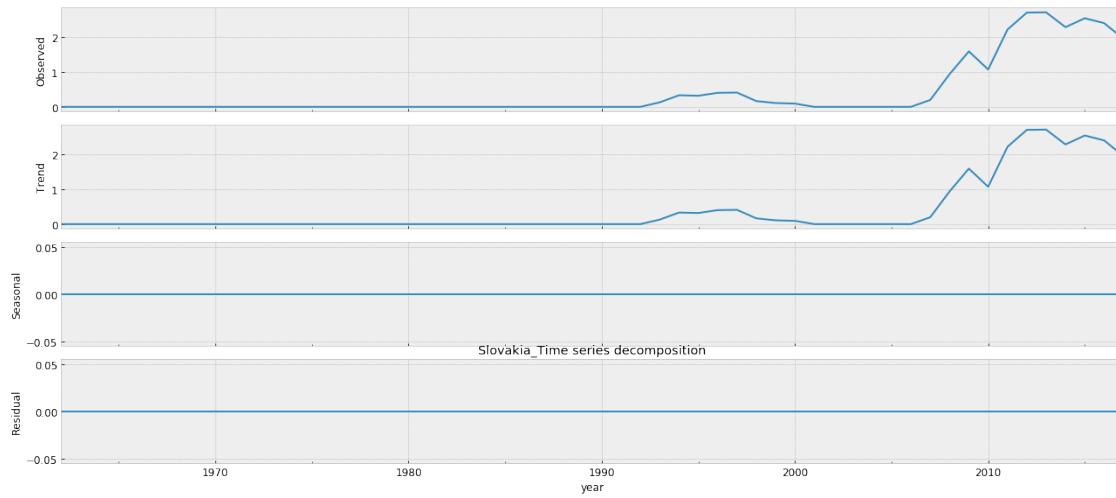
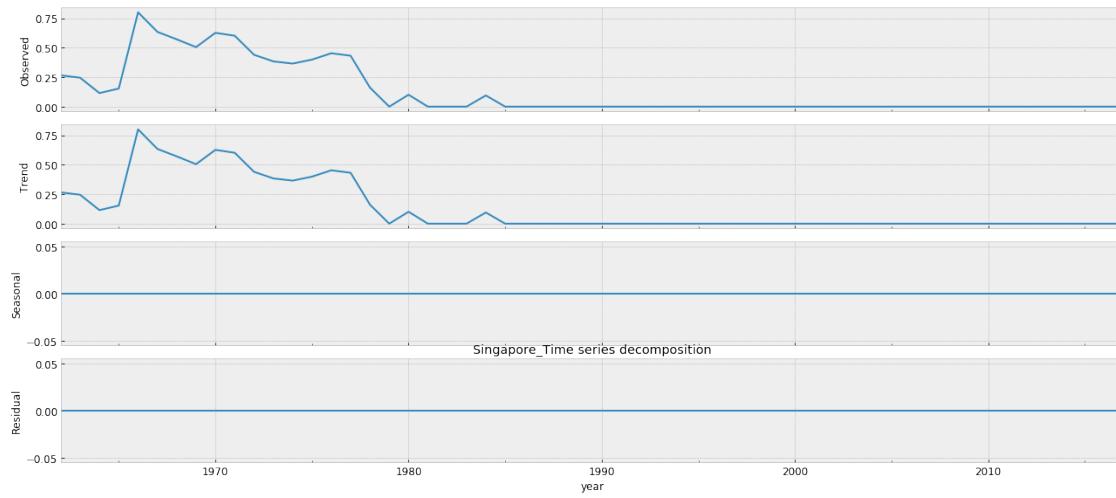
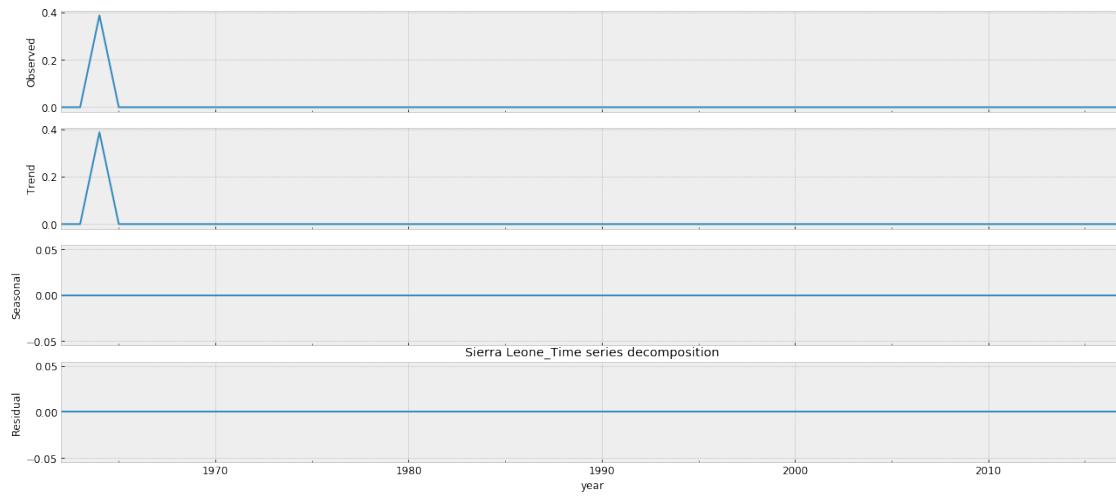


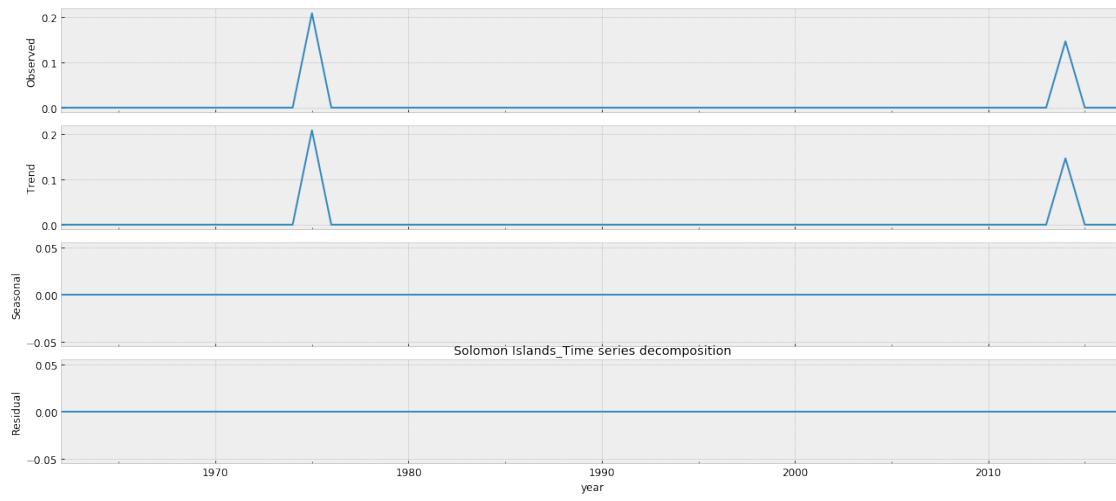
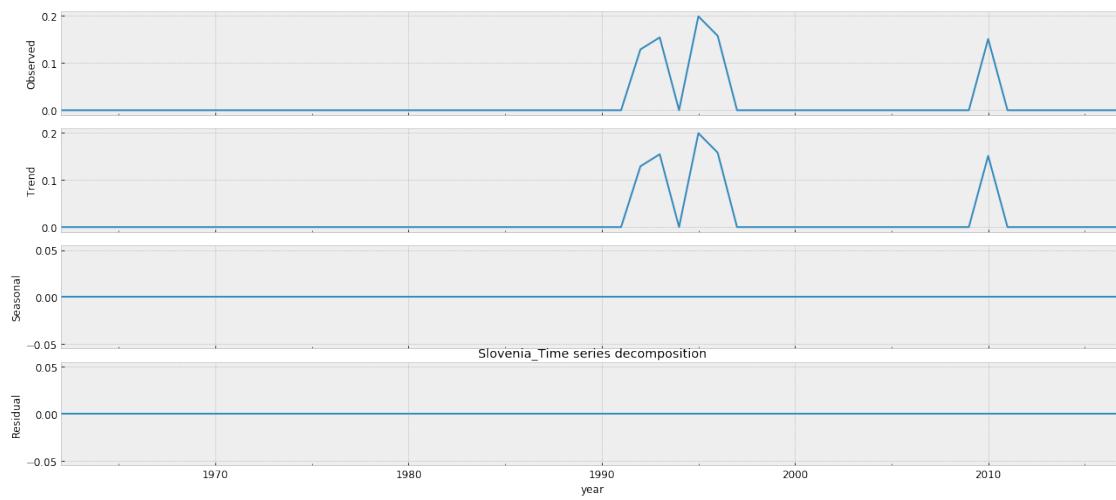


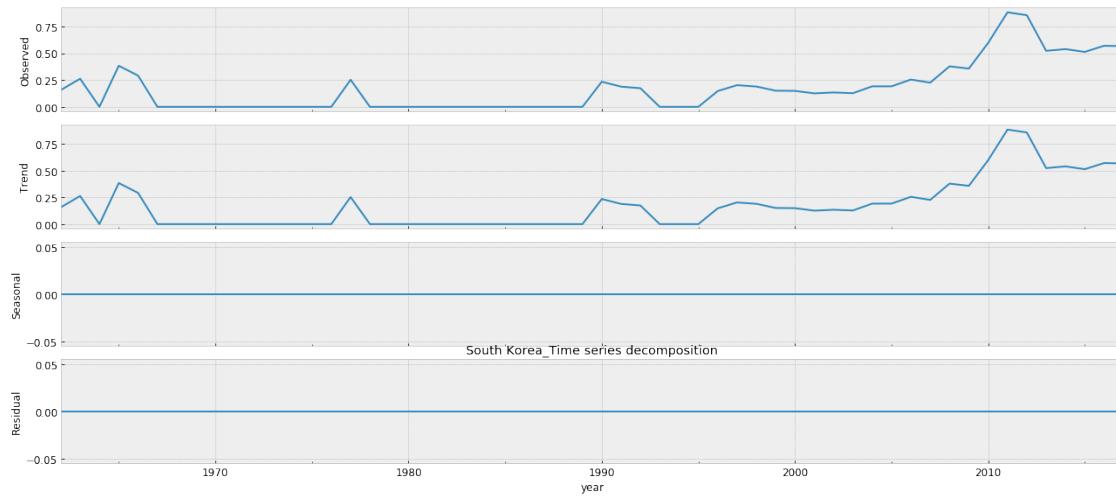
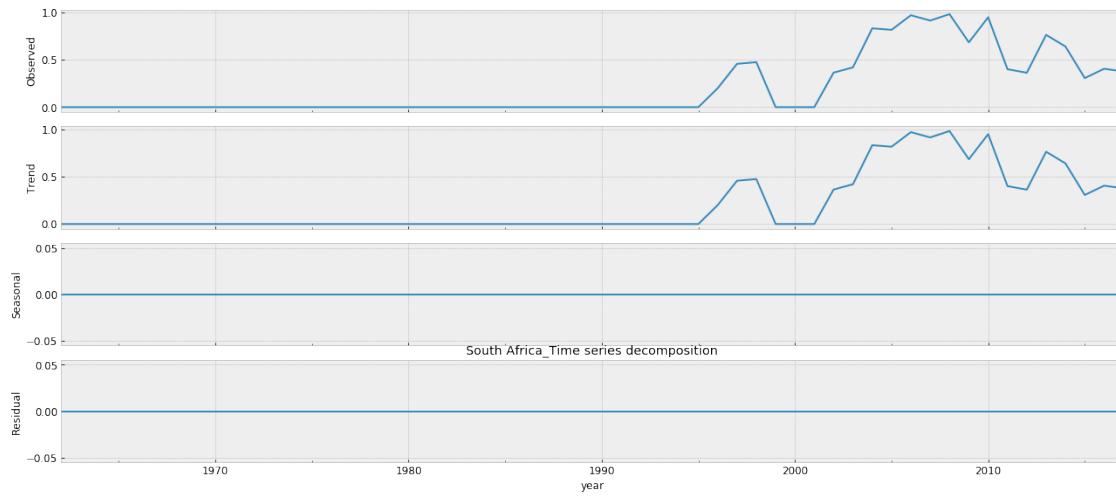
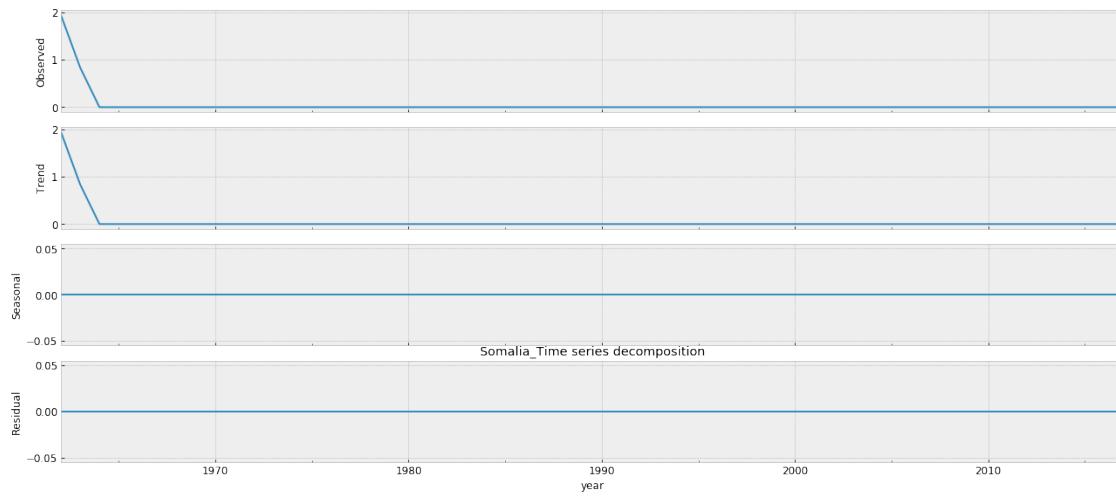


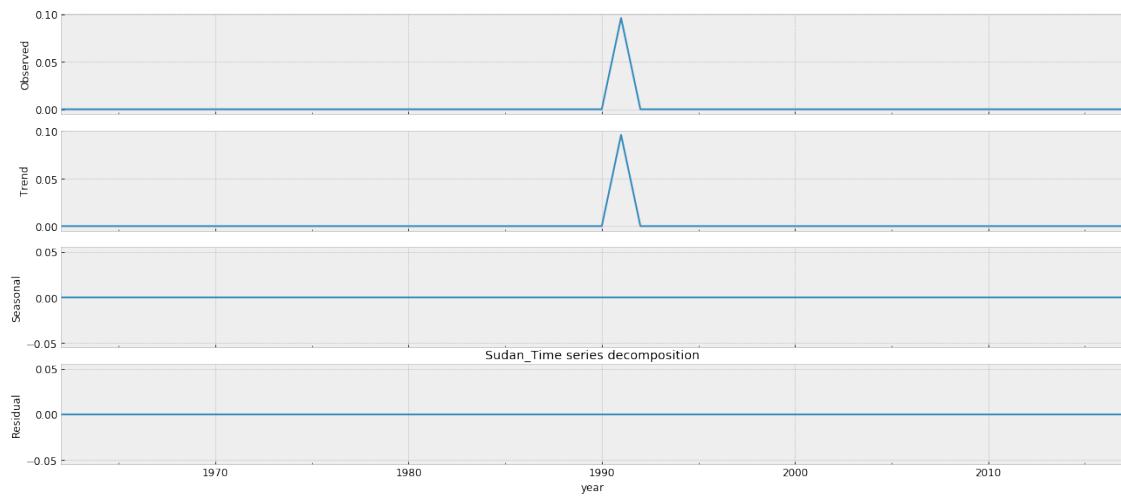
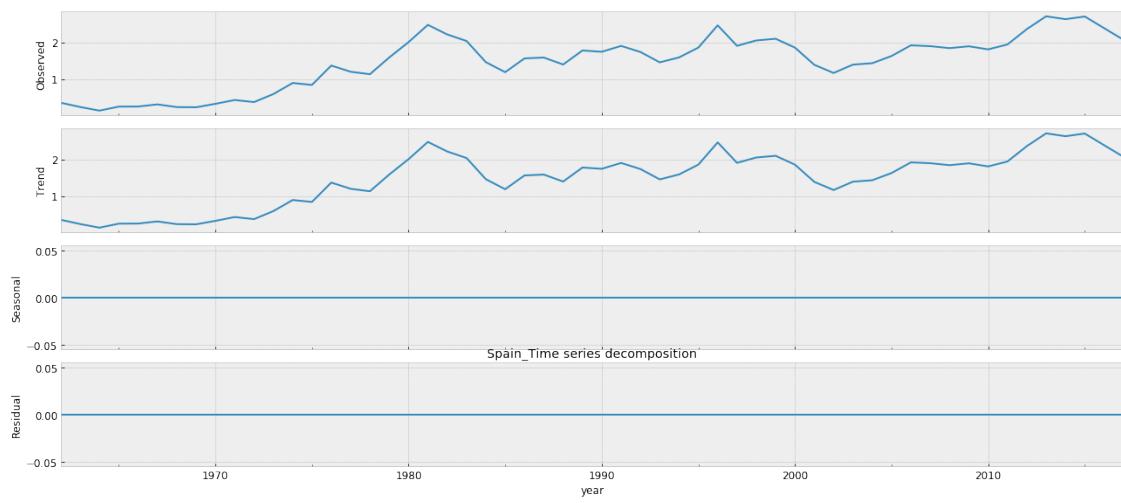


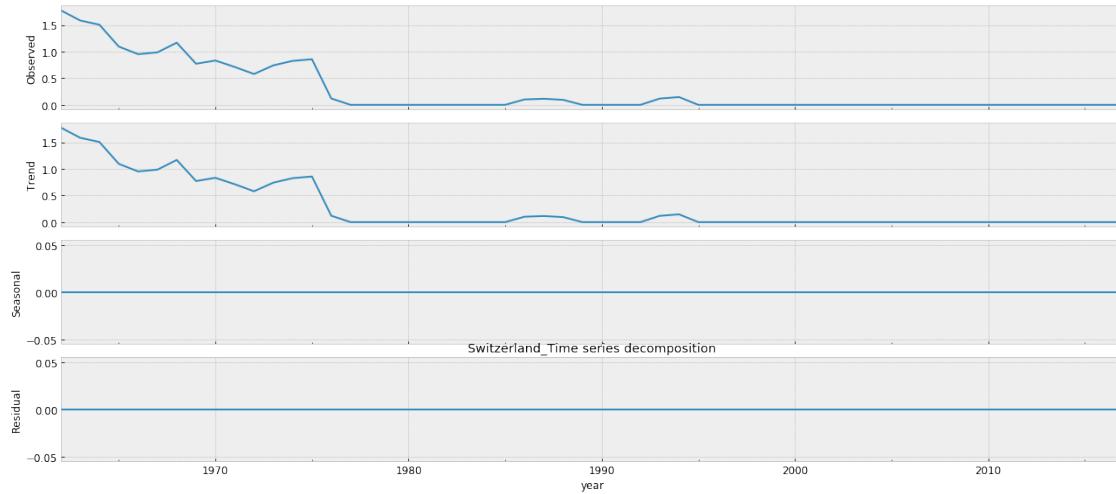
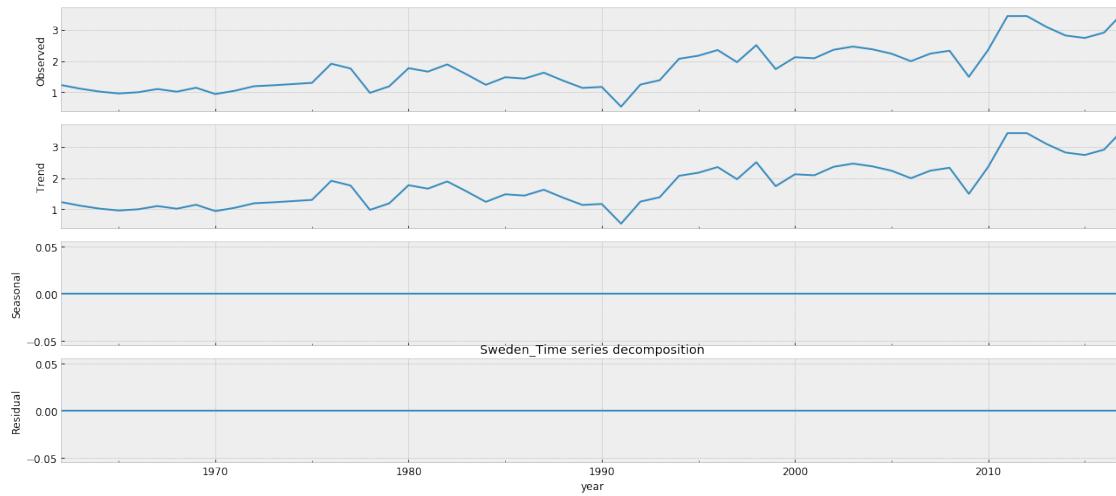
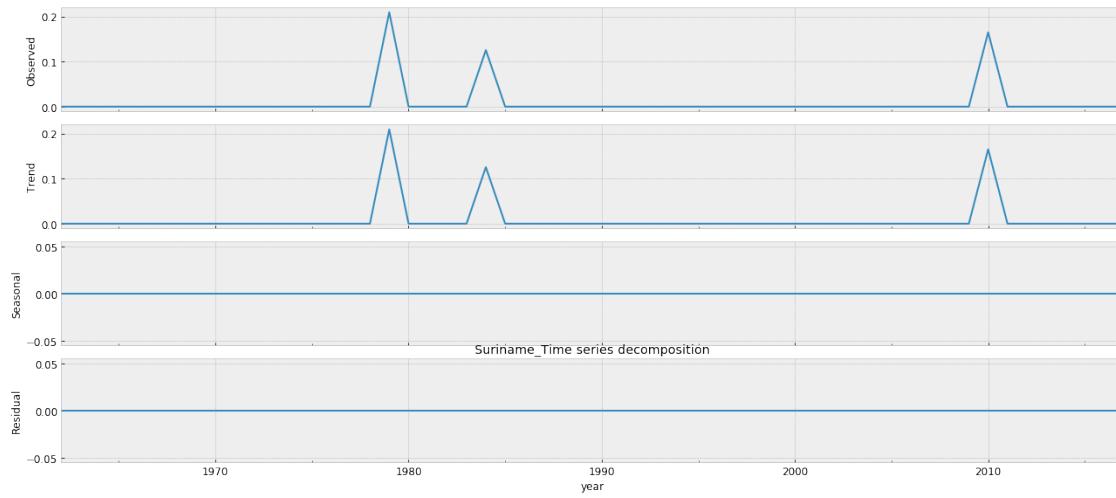


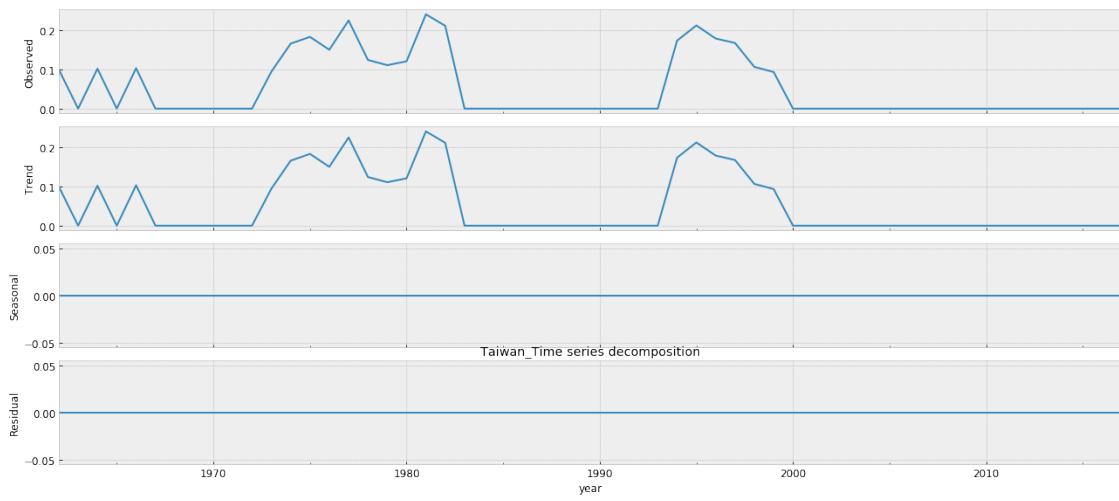
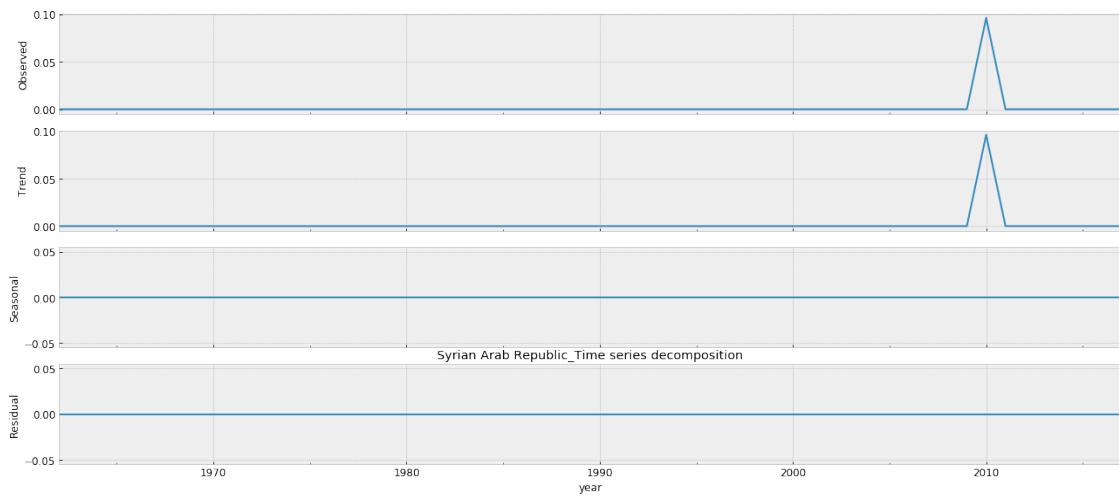


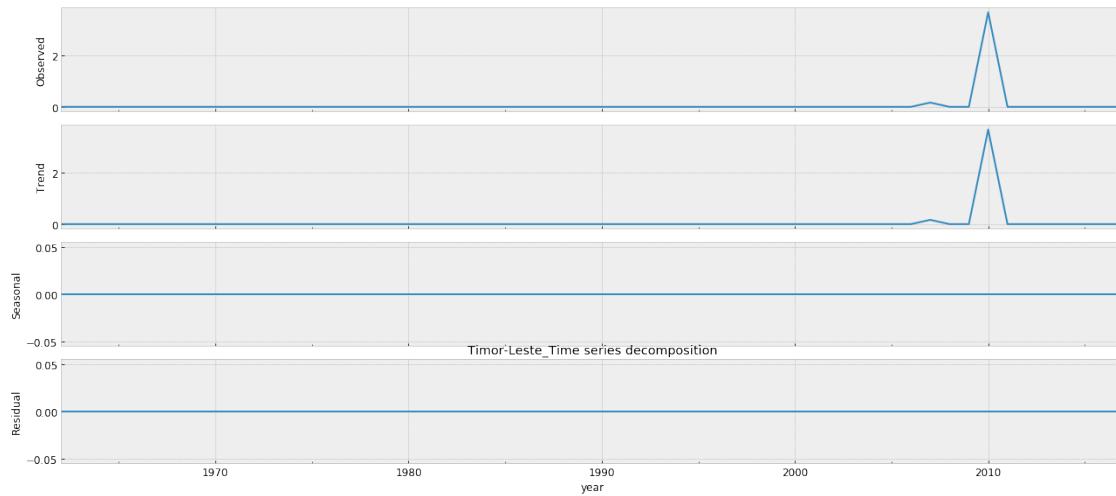
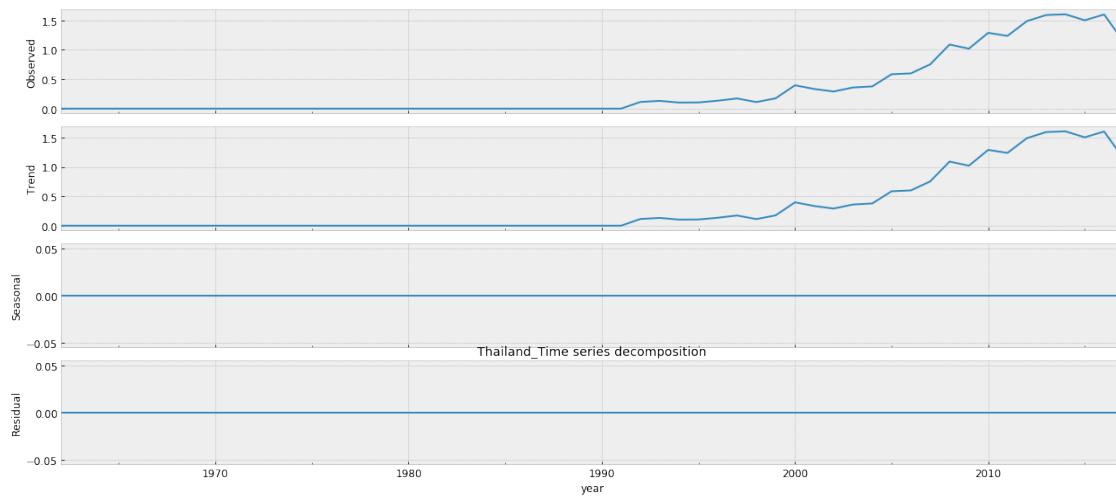
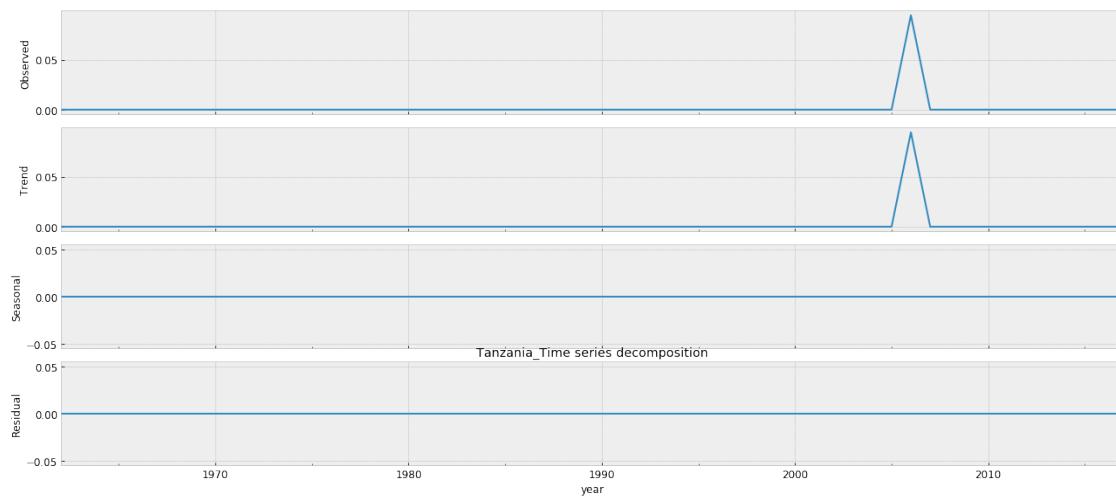


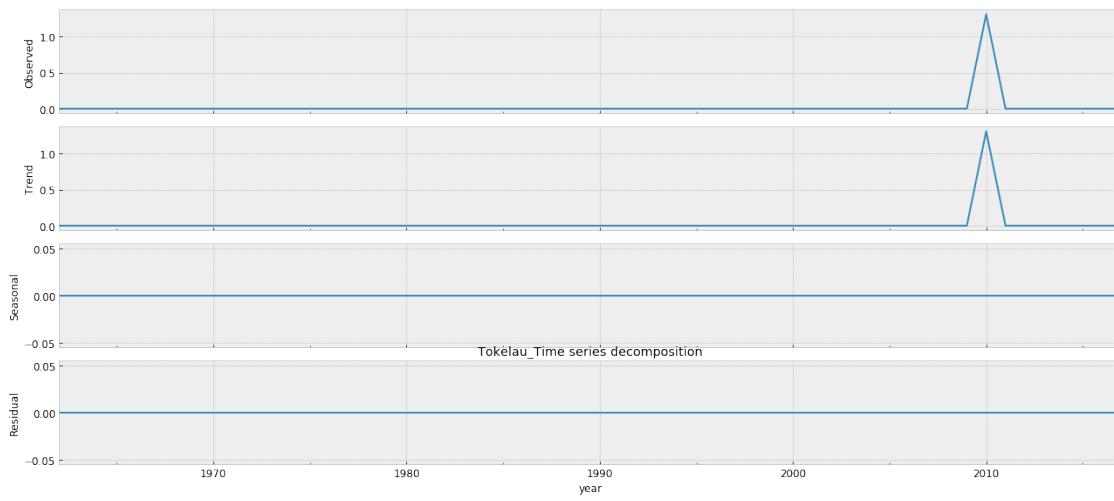
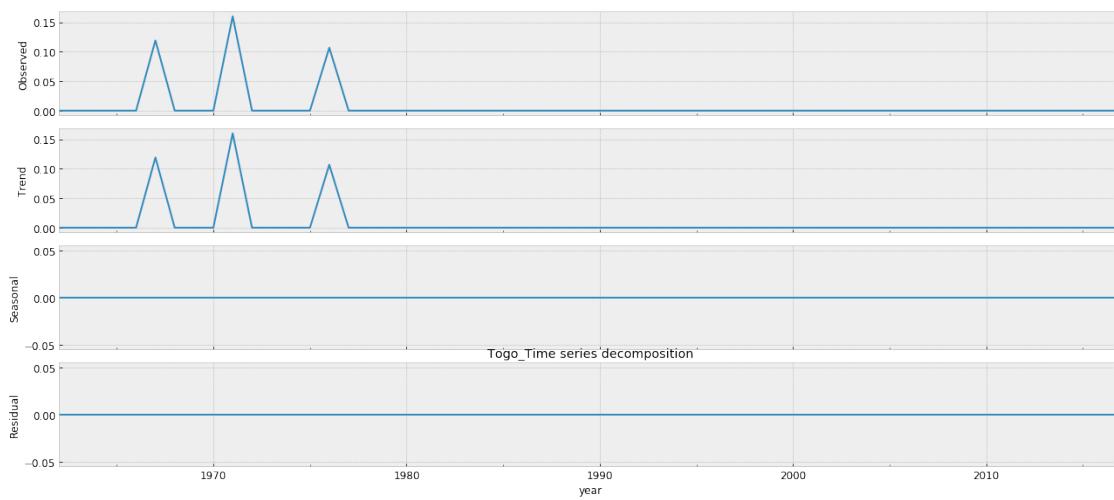


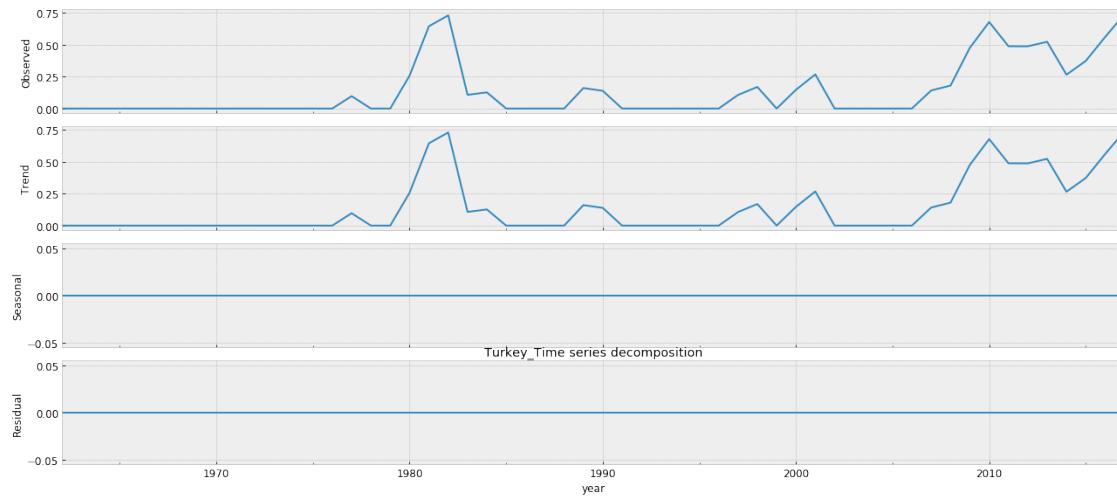
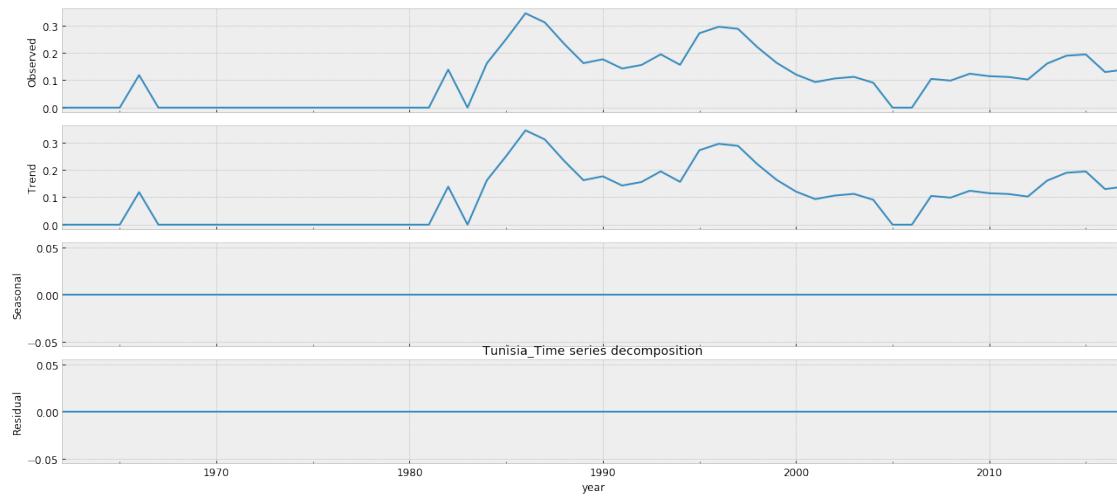
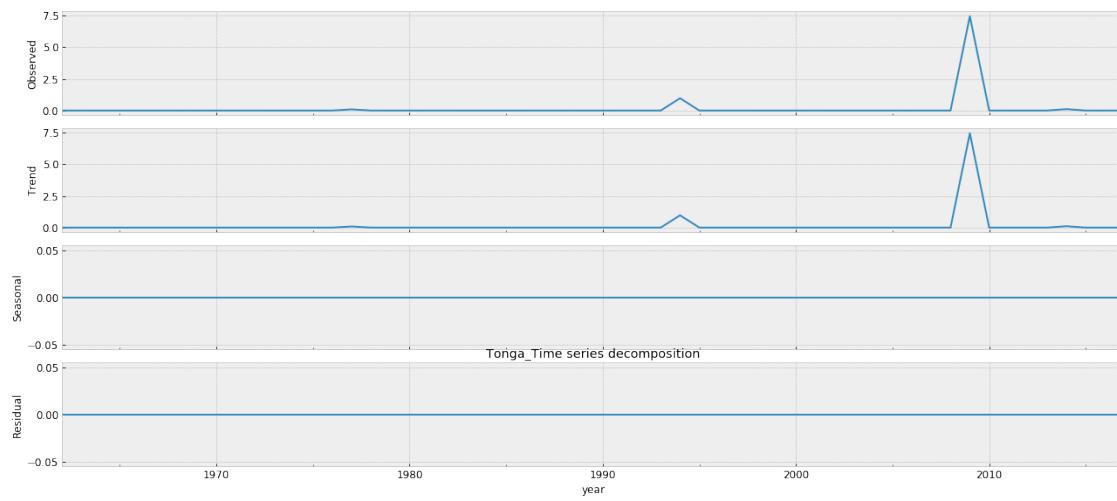


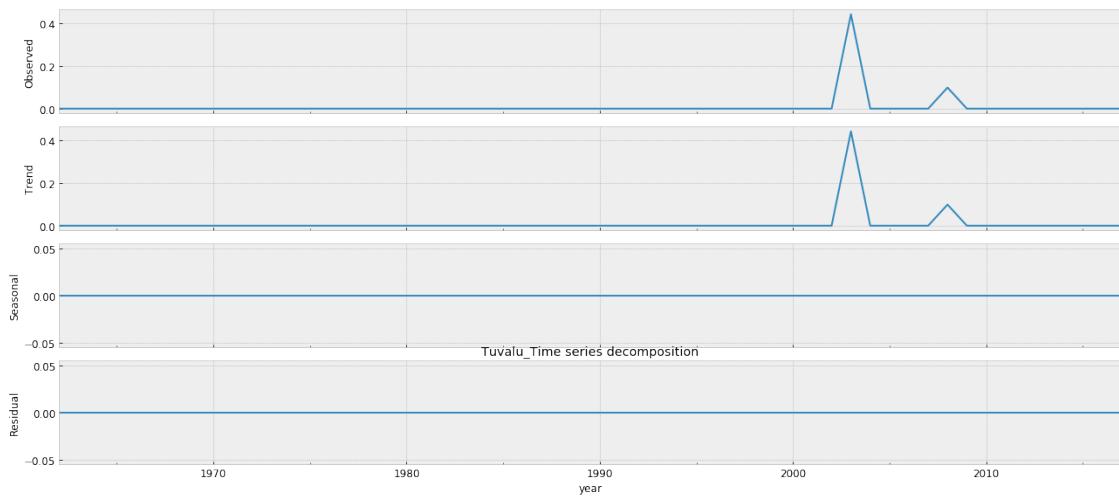
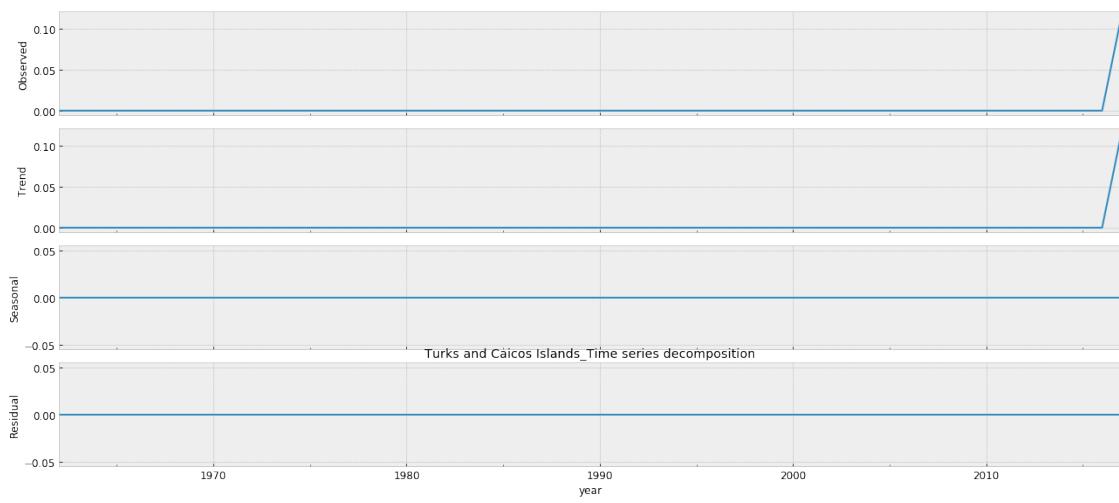


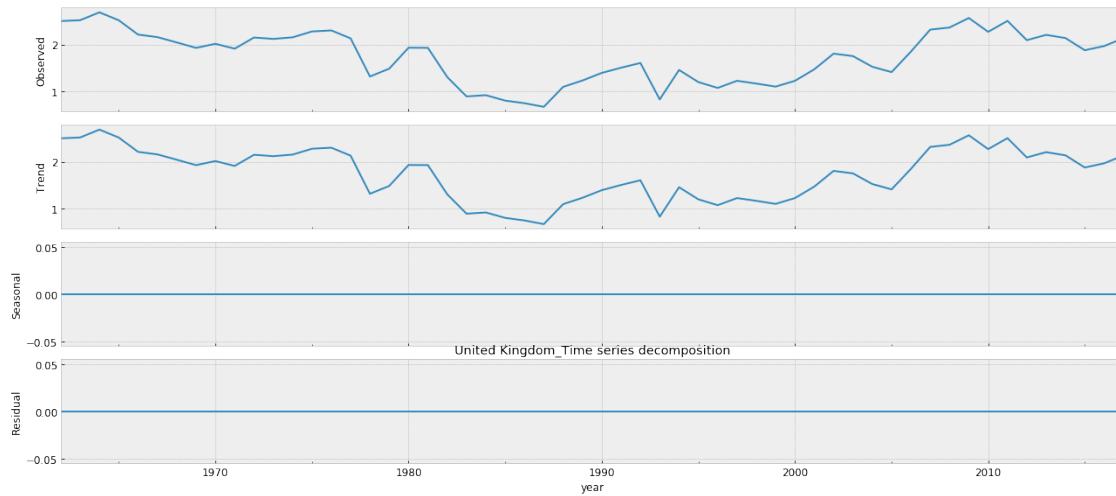
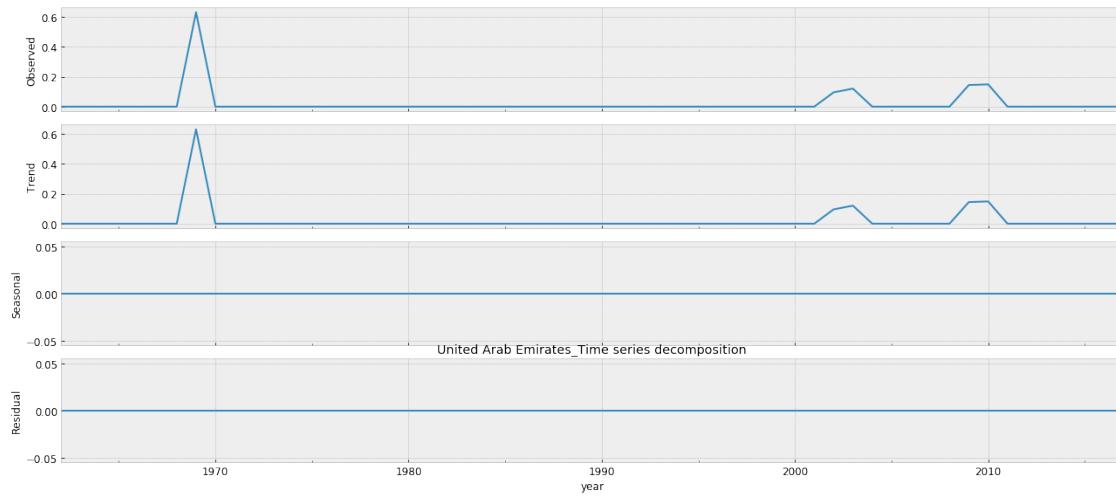
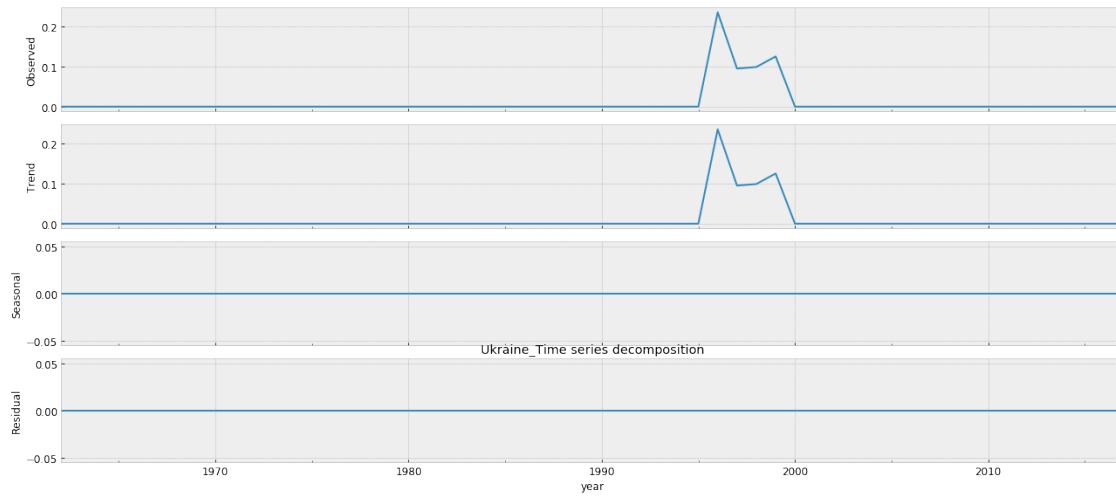


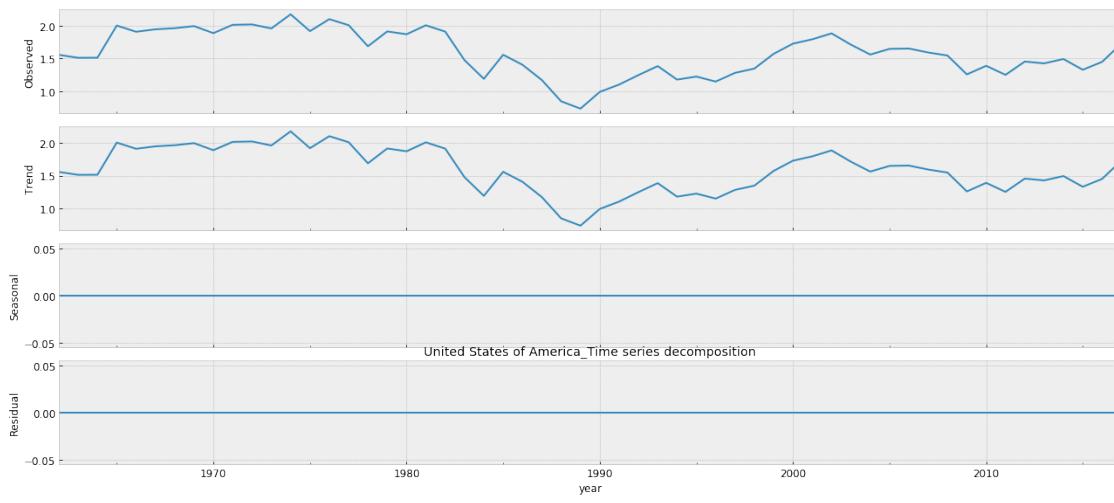
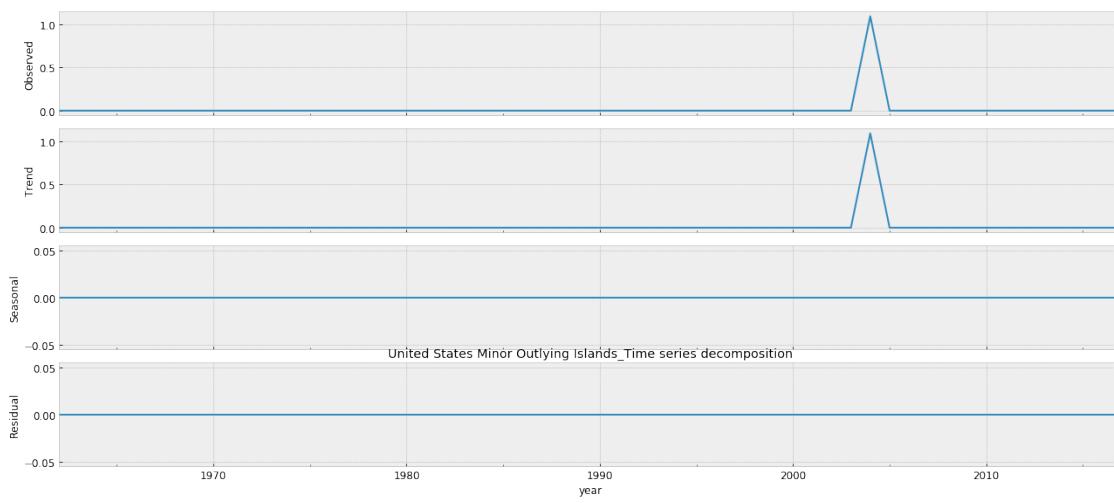


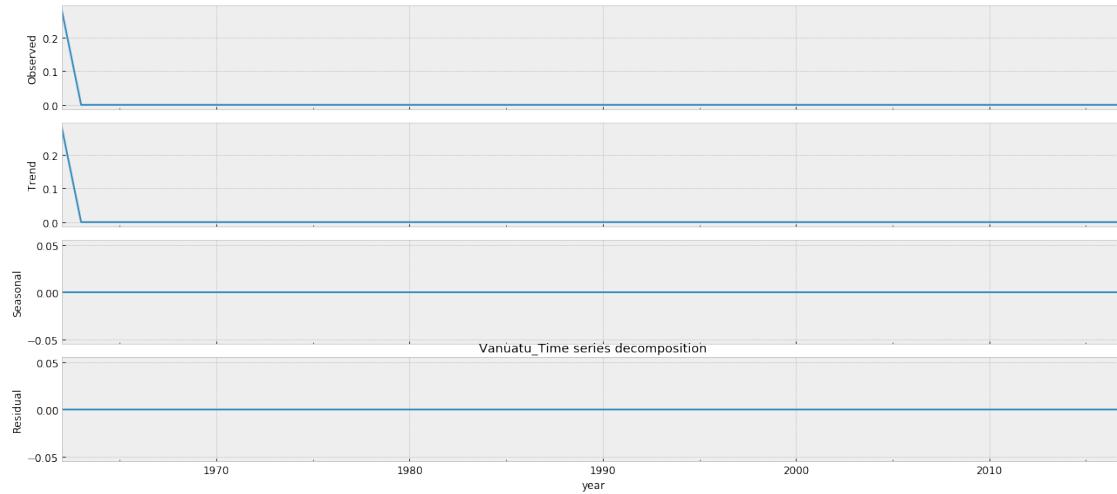
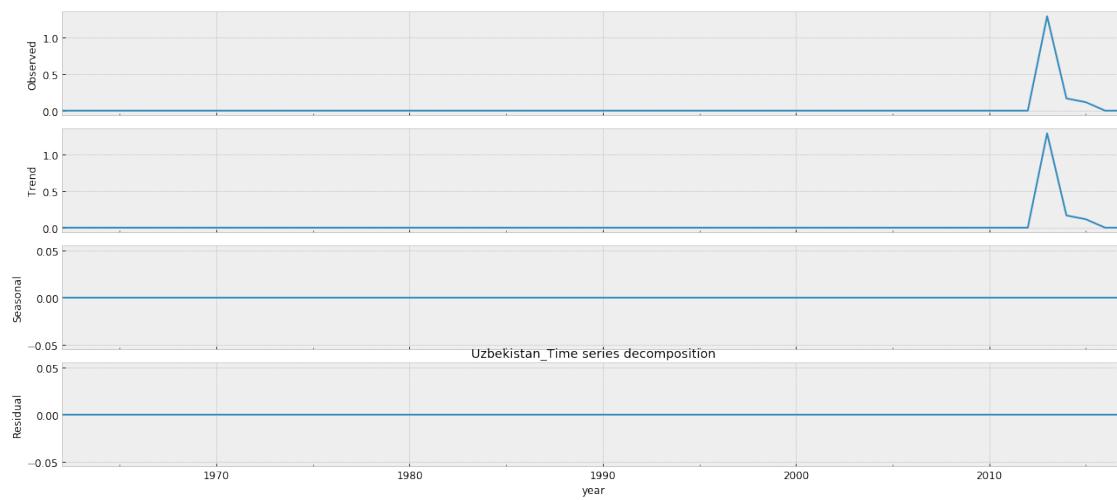
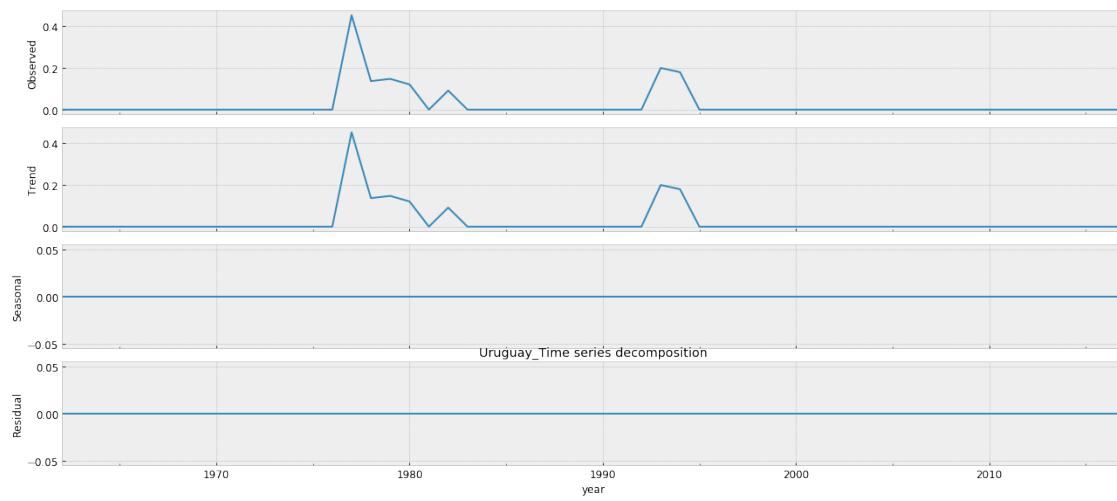


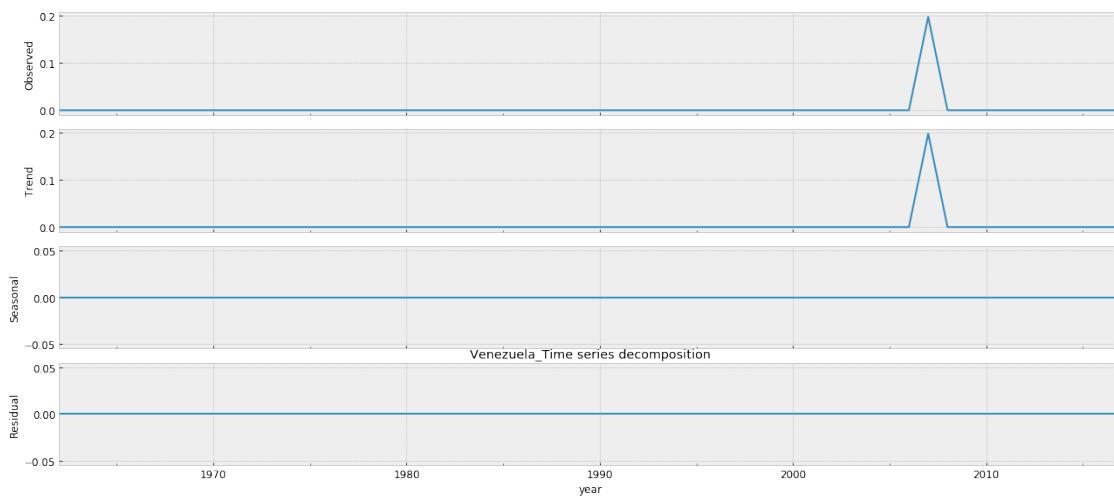
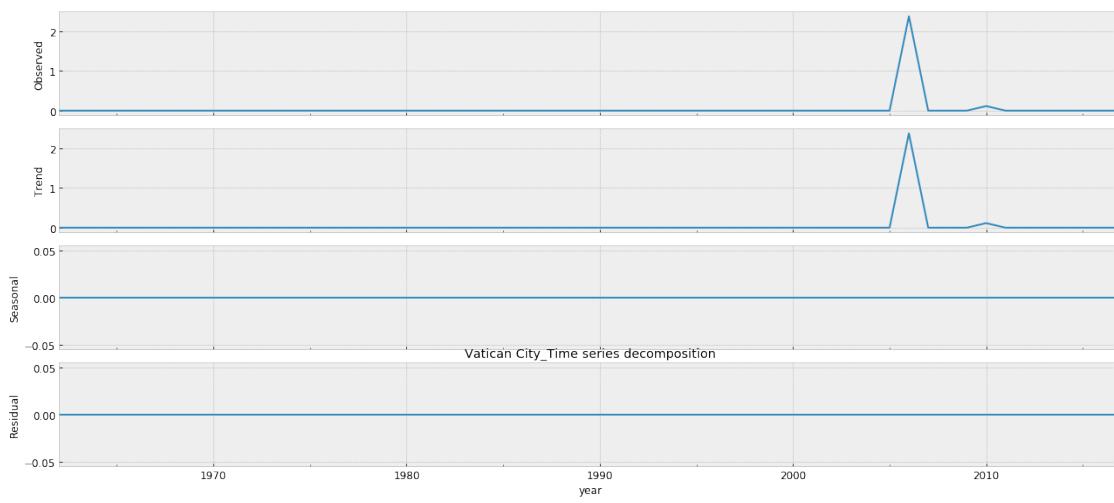


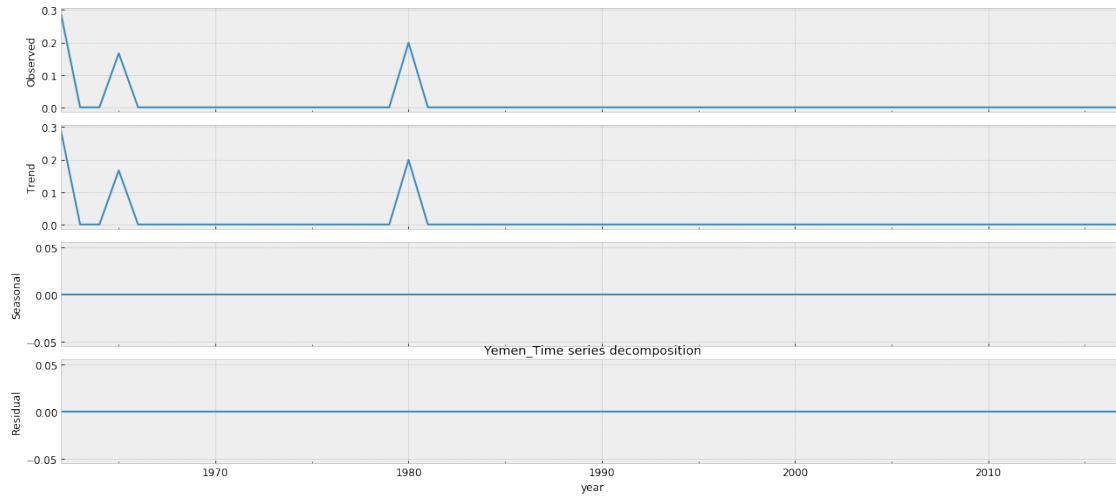
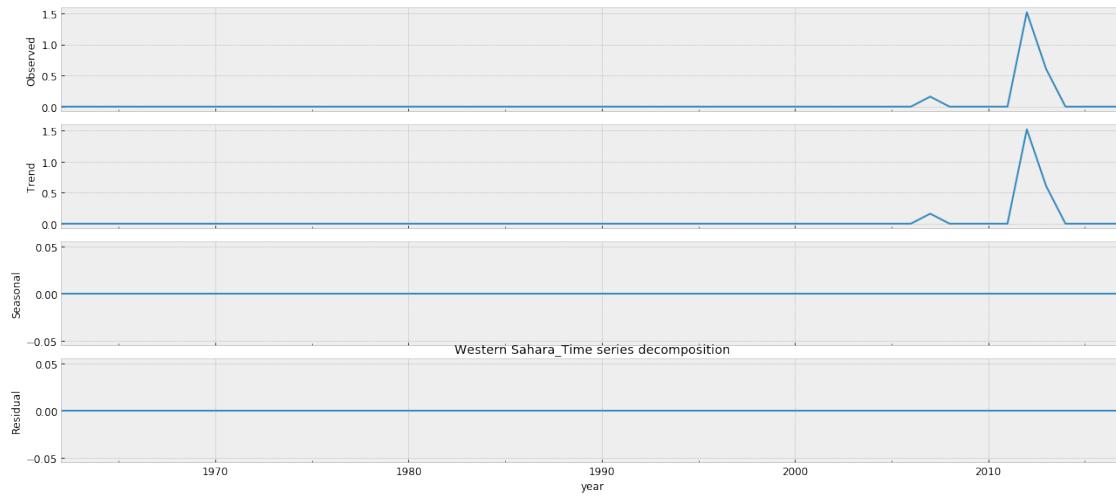
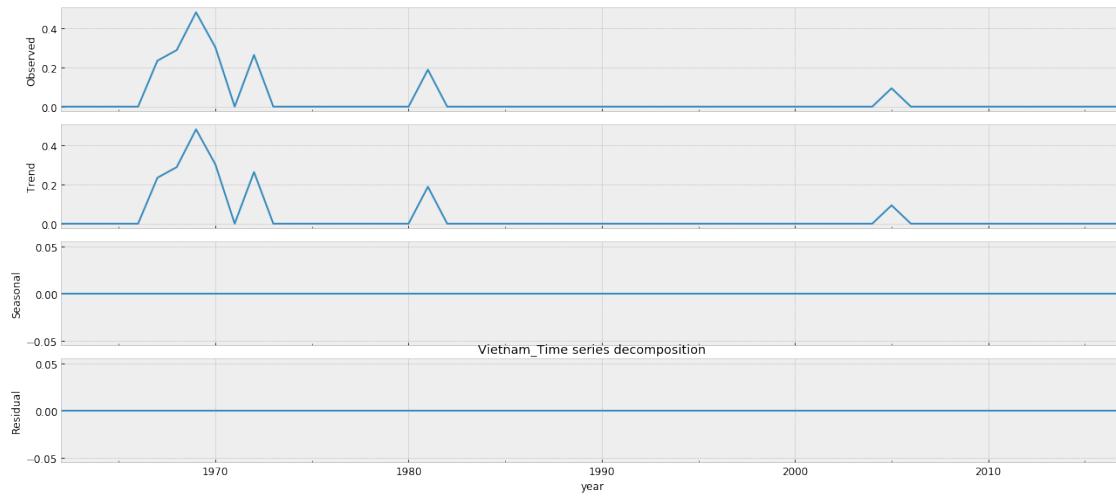


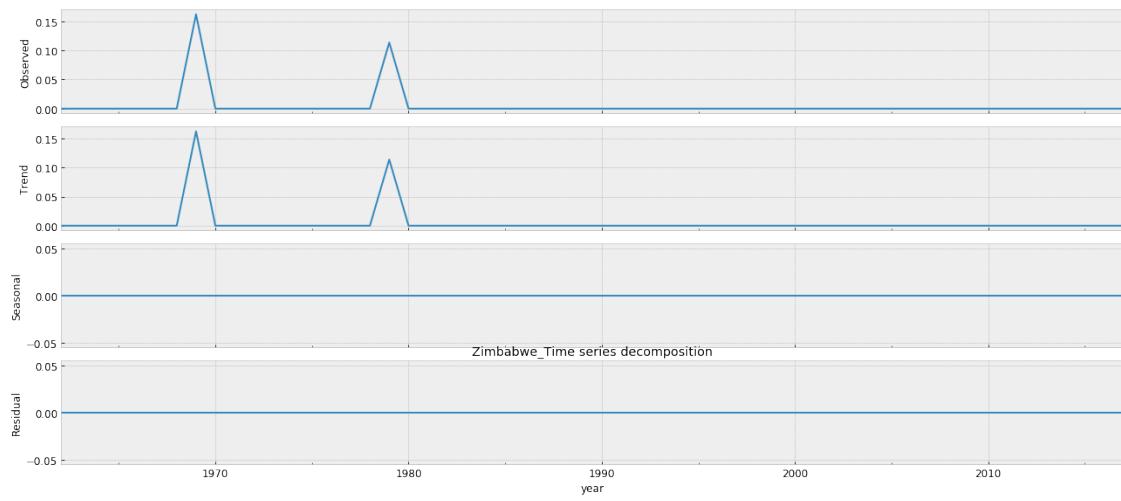
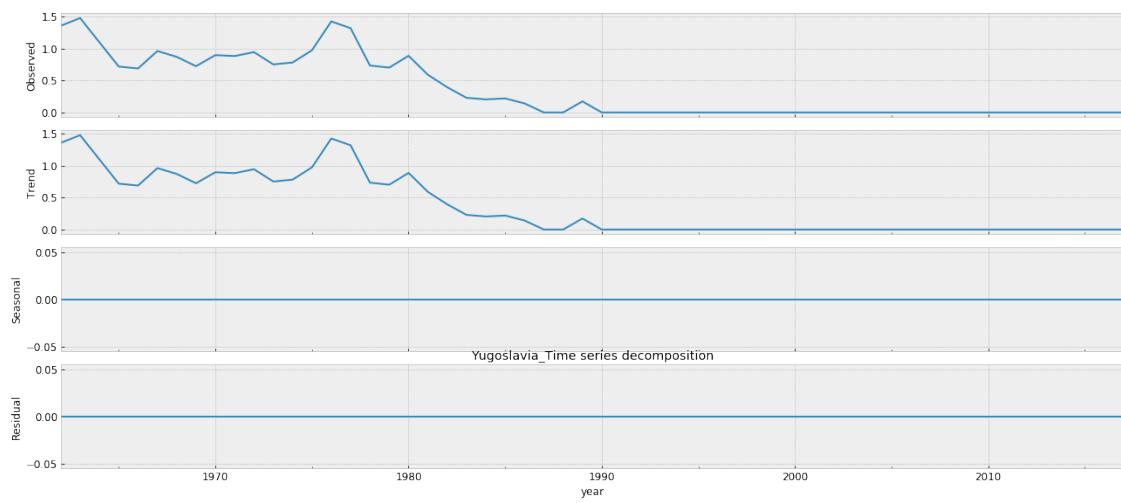


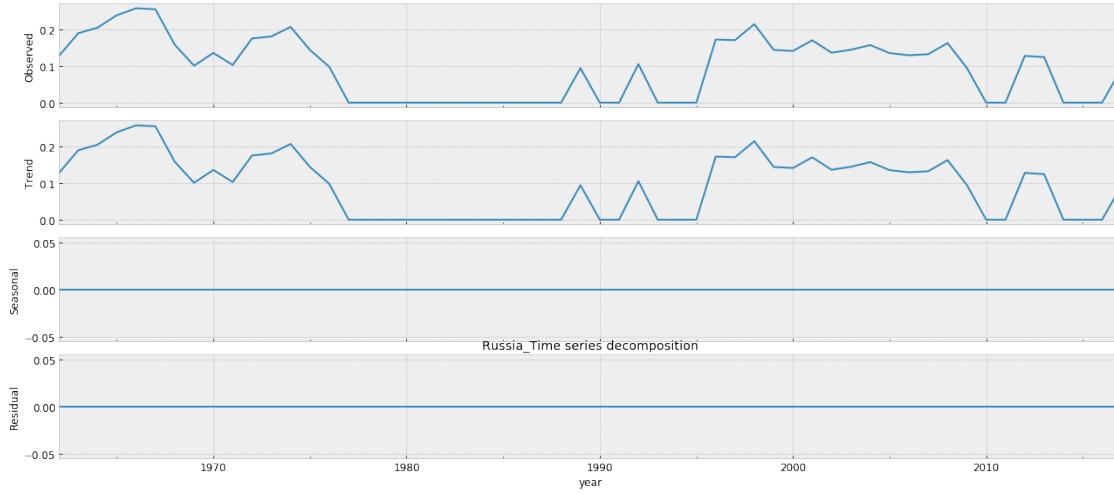












3.1 Stationarity

- Most of the Time Series models work on the assumption that the TS is stationary.
- Major reason for this is that there are many ways in which a series can be non-stationary, but only one way for stationarity.
- Intuitively, we can say that if a Time Series has a particular behaviour over time, there is a very high probability that it will follow the same behaviour in the future.
- Also, the theories related to stationary series are more mature and easier to implement as compared to non-stationary series.

Notes from Alkaline ML

- tests of stationarity for testing a null hypothesis that an observable univariate time series is stationary around a deterministic trend (i.e. trend-stationary).
- A time series is **stationary when its mean, variance and auto-correlation, etc., are constant over time.**
- Many time-series methods may perform better when a time-series is stationary, since forecasting values becomes a far easier task for a stationary time series (high probability of behaving the same way)
- ARIMAs that include differencing (i.e., $d > 0$) assume that the data becomes stationary after differencing.
- This is called **difference-stationary**
- Auto-correlation plots are an easy way to determine whether your time series is sufficiently stationary for modeling.
- If the plot does not appear relatively stationary, your model will likely need a differencing term.
- These can be determined by using an Augmented Dickey-Fuller test, or various other statistical testing methods.
- Note that `auto_arima` will automatically determine the appropriate differencing term for you by default.

In order to quantitatively determine whether we need to difference our data in order to make it

stationary, we can conduct a test of stationarity

We can check stationarity using the following:

1. **ACF and PACF plots:**

- If the time series is stationary, the ACF/PACF plots will show a **quick drop-off in correlation** after a small amount of lag between points.

2. **Plotting Rolling Statistics:**

- We can plot the moving average or moving variance and see if it varies with time.
- **Moving average/variance** is for any instant ‘t’, the average/variance of the last period (e.g year or last 12 months)

3. **Augmented Dickey-Fuller Test [*]:**

- This is one of the statistical tests for checking stationarity.
- Here the **null hypothesis is that the TS is non-stationary**.
- The null hypothesis of the Augmented Dickey-Fuller is that there is a unit root, with the alternative that there is no unit root.
- If the pvalue is above a critical size, then we cannot reject that there is a unit root.
- The p-values are obtained through regression surface approximation from MacKinnon 1994, but using the updated 2010 tables
- If the p-value is close to significant, then the critical values should be used to judge whether to reject the null.
- The autolag option and maxlag for it are described in Greene
- The test results comprise of a Test Statistic and some Critical Values for difference confidence levels.
- If the ‘Test Statistic’ is less than the ‘Critical Value’, we can reject the null hypothesis and say that the series is stationary.

References:

1. W. Green. “Econometric Analysis,” 5th ed., Pearson, 2003.
2. Hamilton, J.D. “Time Series Analysis”. Princeton, 1994.
3. MacKinnon, J.G. 1994. ”Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business and Economic Statistics* 12, 167-76.
4. MacKinnon, J.G. 2010. “Critical Values for Cointegration Tests.” Queen’s University, Dept of Economics, Working Papers. Available at <http://ideas.repec.org/p/qed/wpaper/1227.html>

3.1.1 Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots

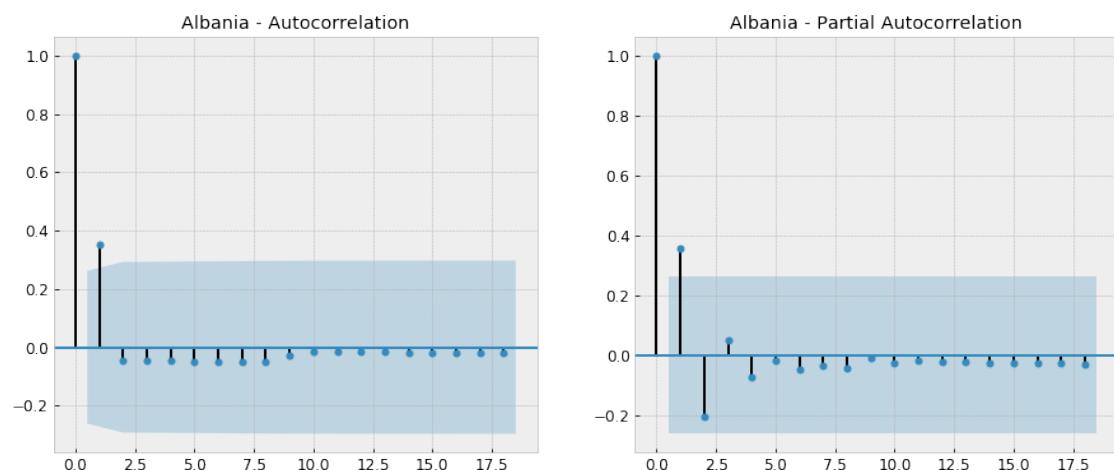
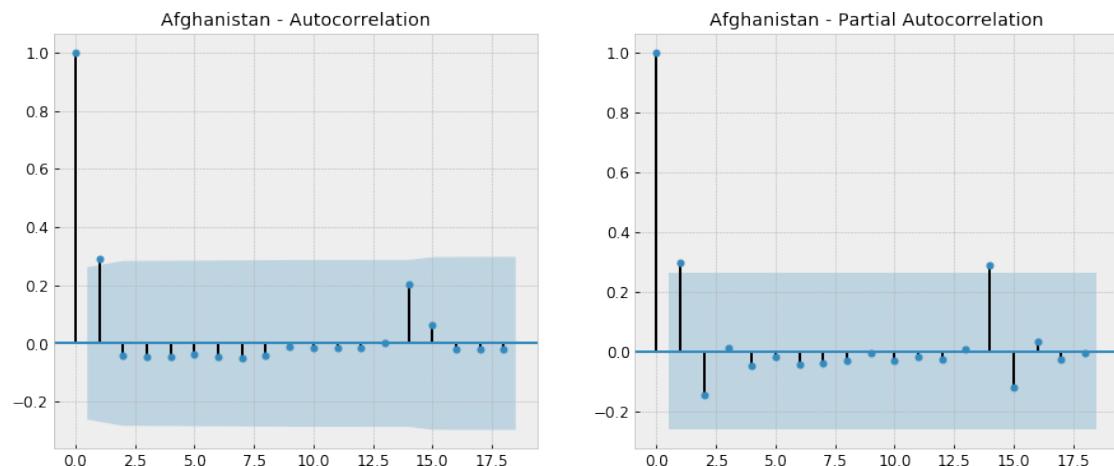
- If the time series is stationary, the ACF/PACF plots will show a **quick drop-off in correlation** after a small amount of lag between points.
- Most of this data is stationary as a high number of previous observations are not correlated with future values.
- Confidence intervals are drawn as a cone.
- By default, this is set to a 95% confidence interval, suggesting that correlation values outside of this cone are very likely a correlation and not a statistical fluke.
- The partial autocorrelation at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags.

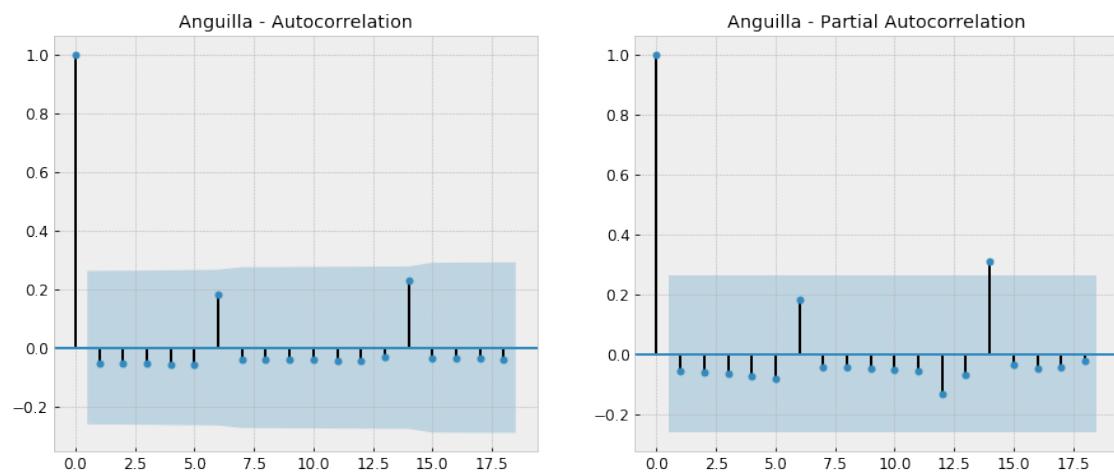
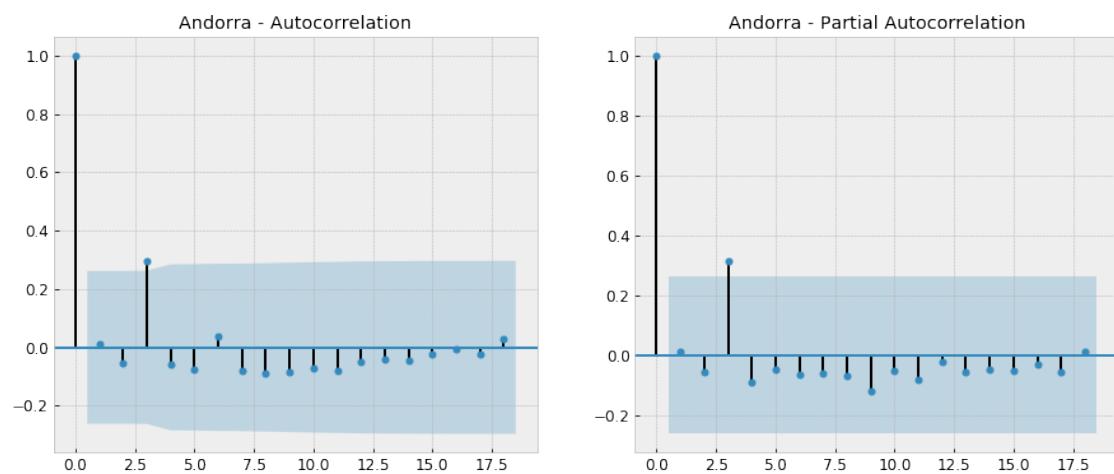
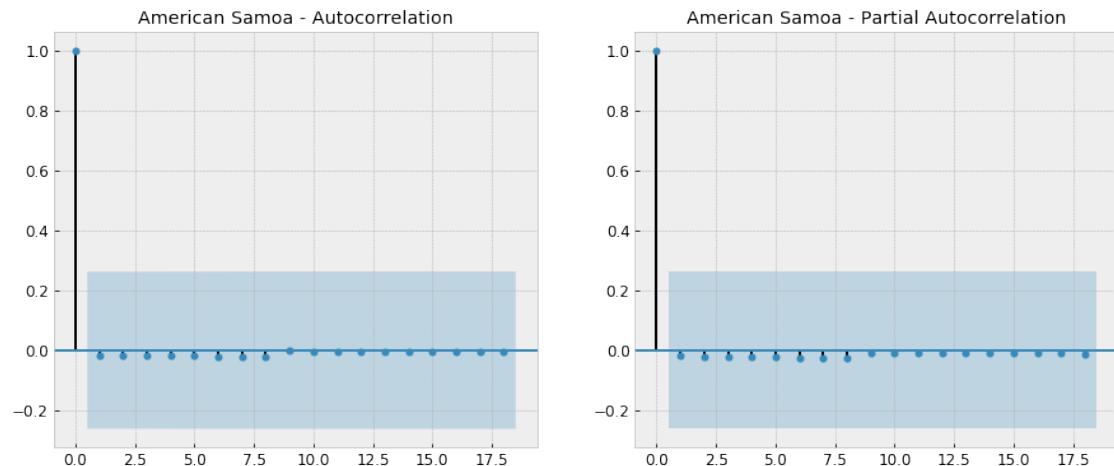
```
[177]: for country in countries:
    savefile = save_images(df_grps, 'acf_pacf')

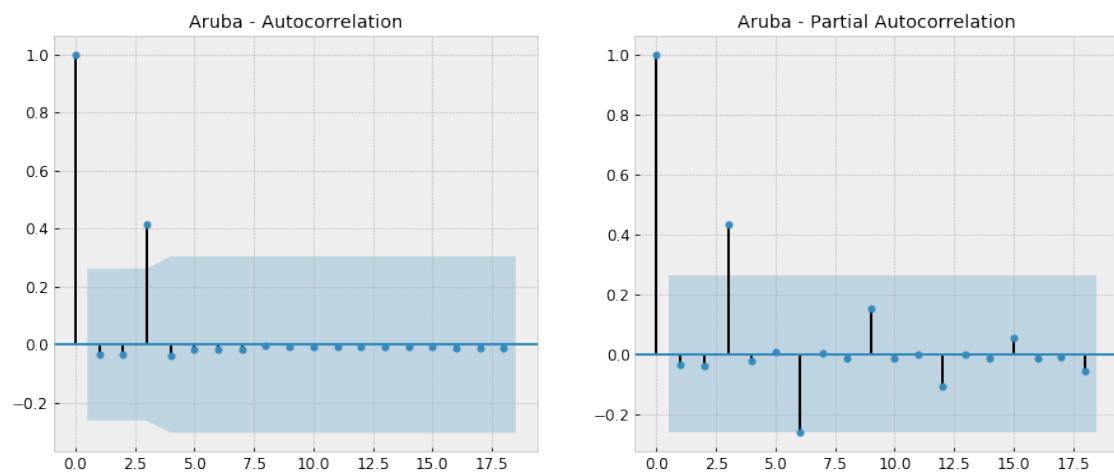
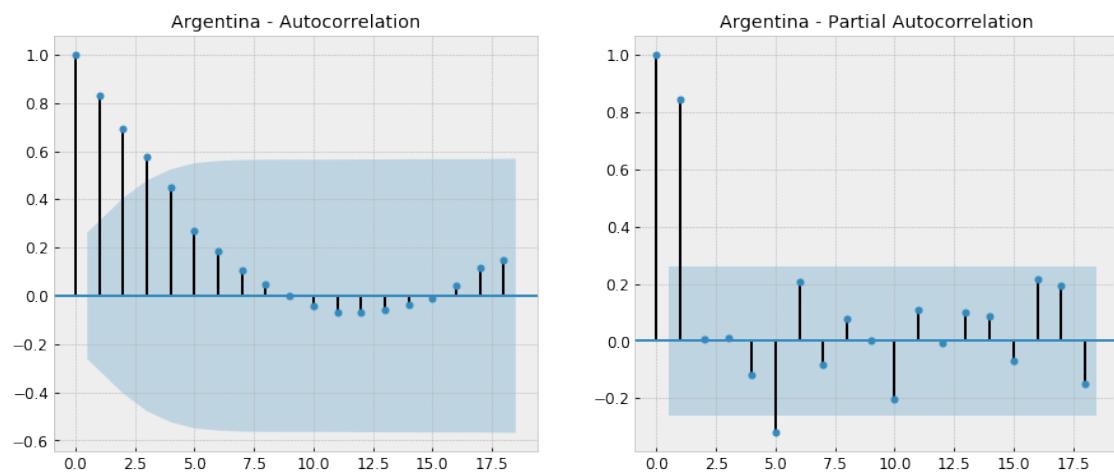
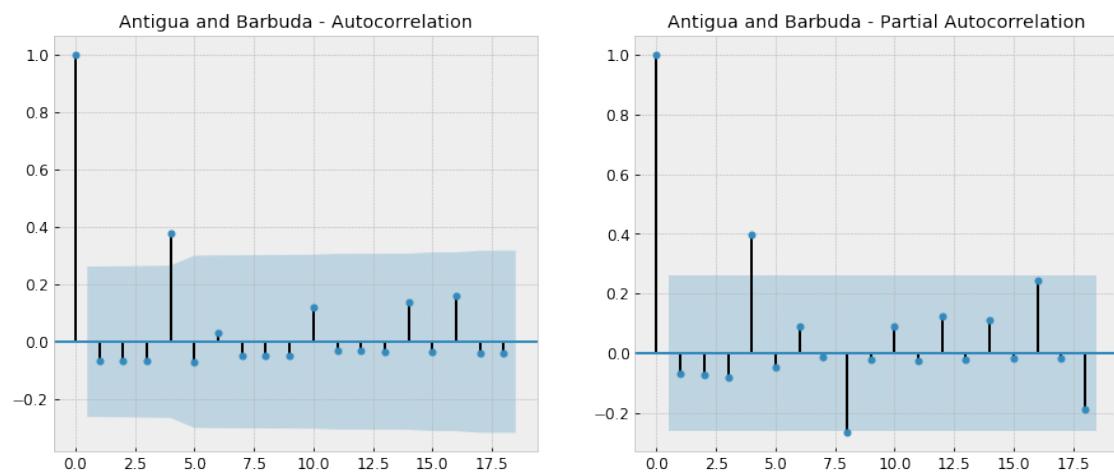
    plt.figure(figsize=(15,6))
    plt.subplot(121)
    plot_acf(df_grps[country], ax=plt.gca(), title=f'{df_grps[country].name} - Autocorrelation')

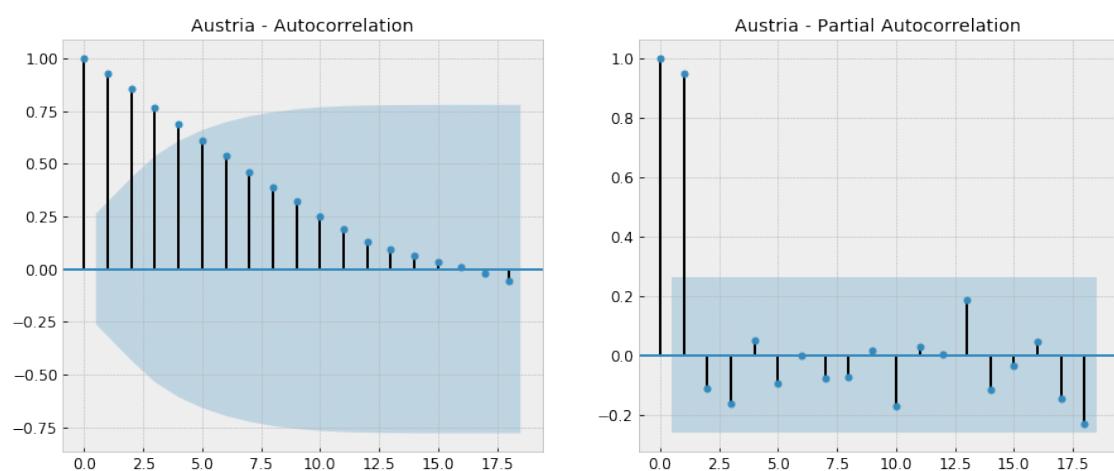
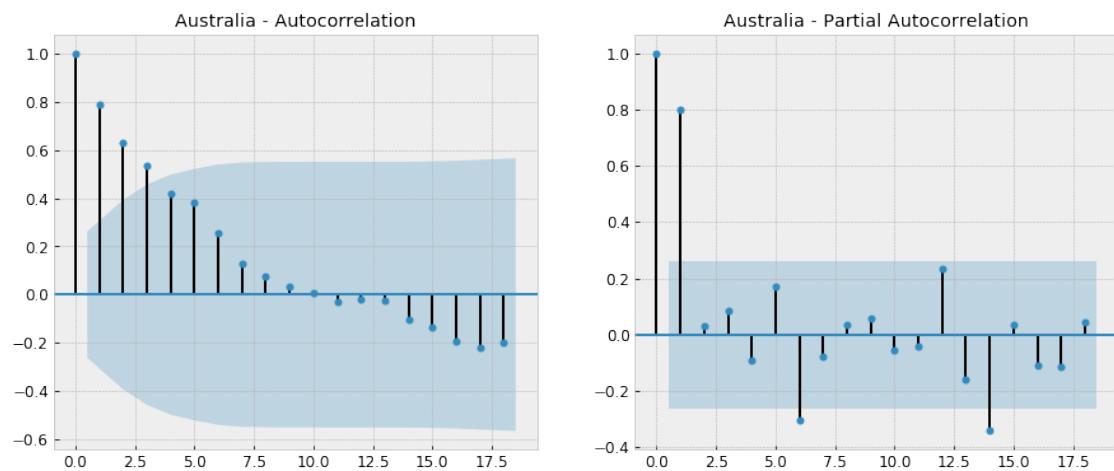
    plt.subplot(122);
    plot_pacf(df_grps[country], ax=plt.gca(), title=f'{df_grps[country].name} - Partial Autocorrelation')

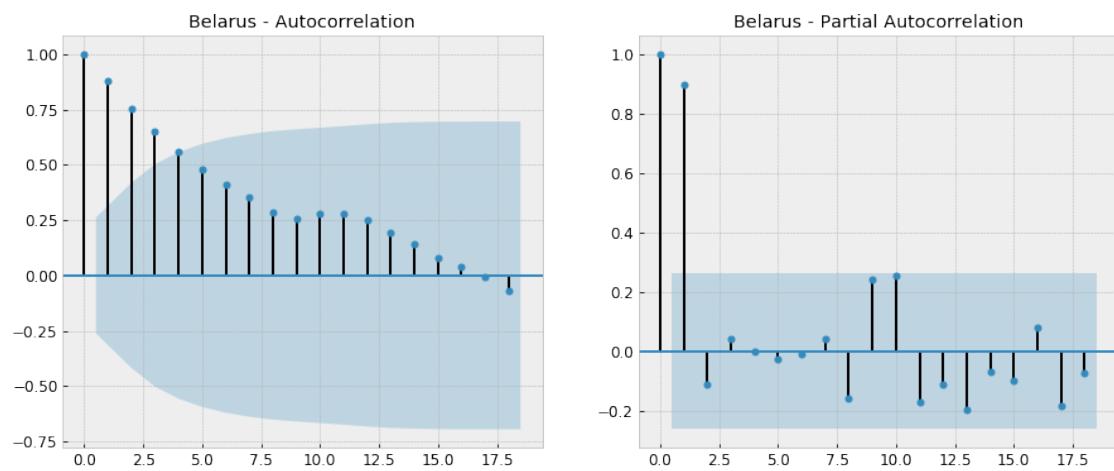
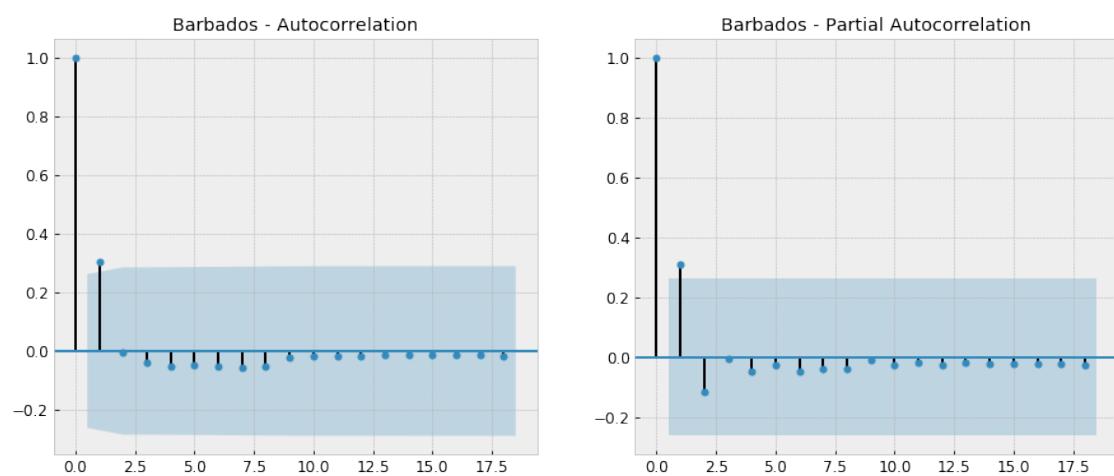
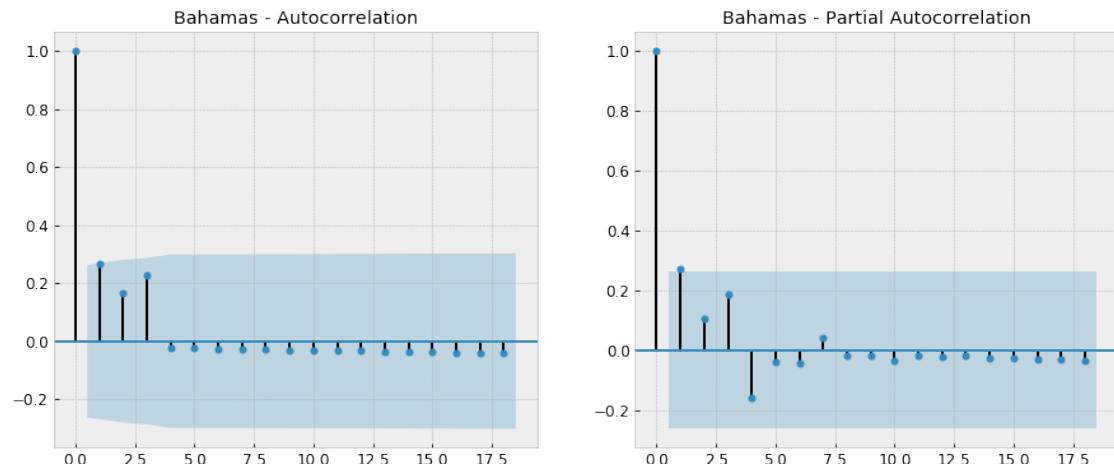
    plt.savefig(str('{}_ACF_PACF'.format(savefile)))
    plt.show()
```

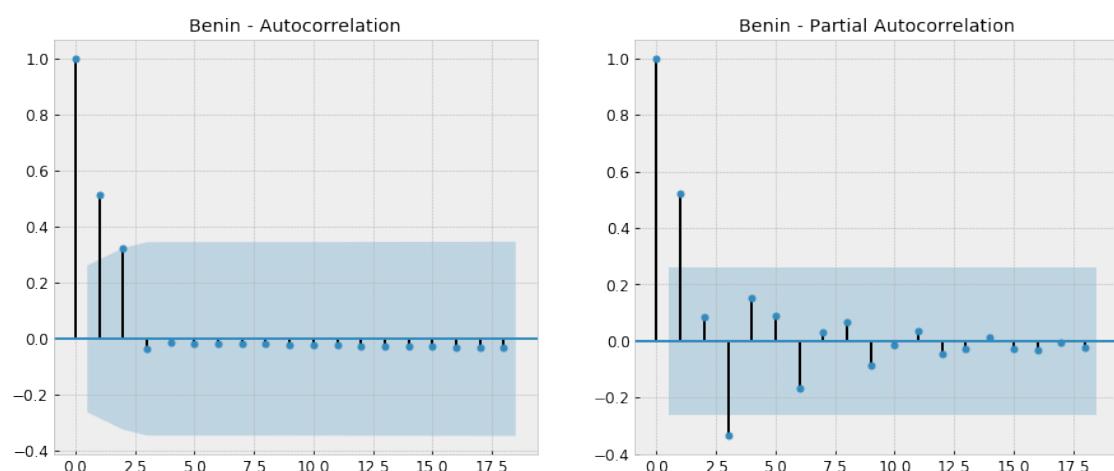
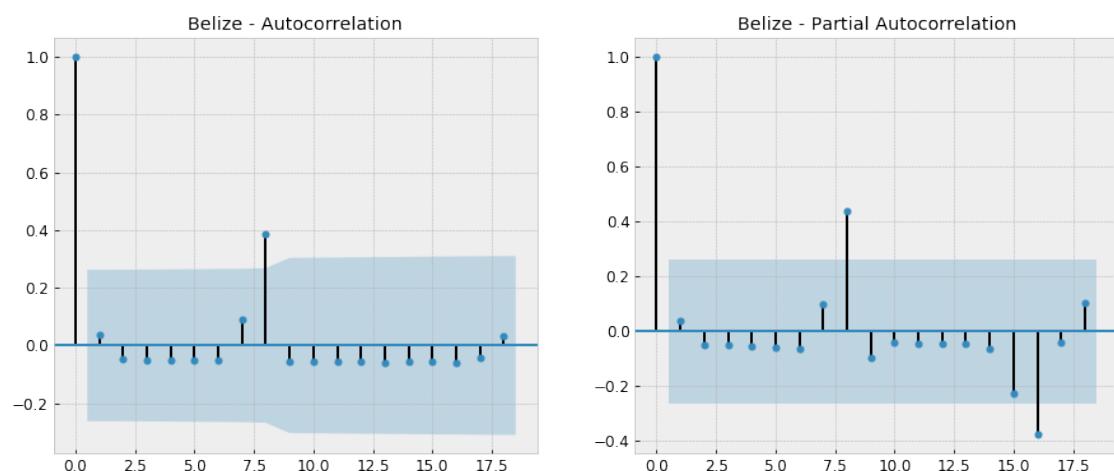
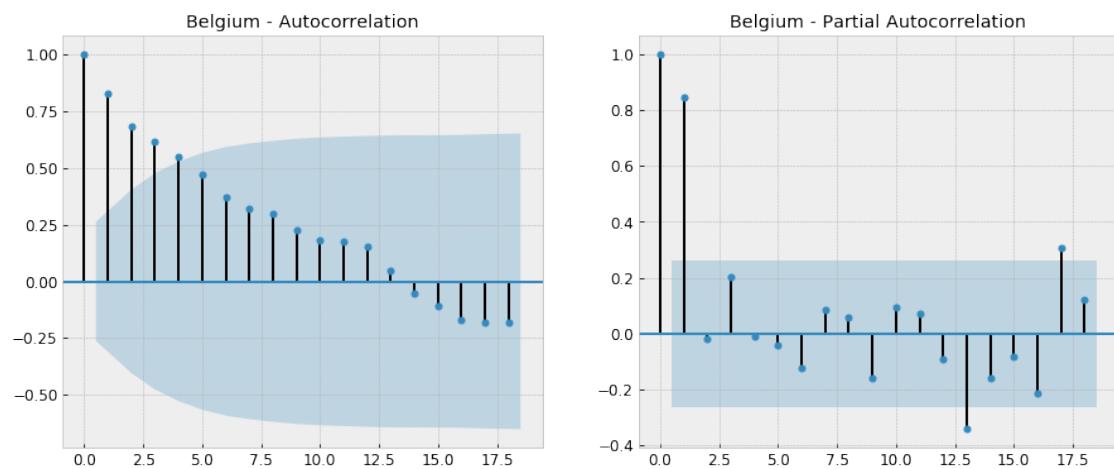


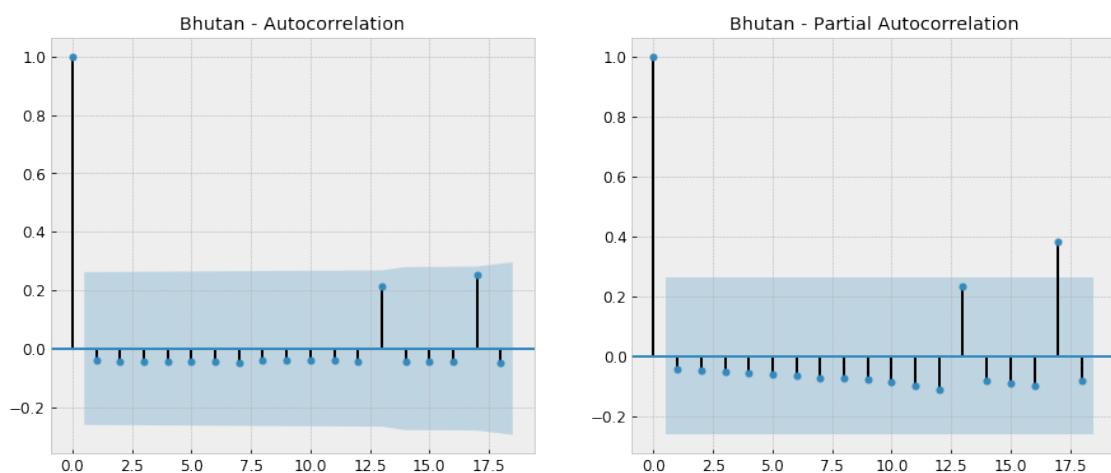
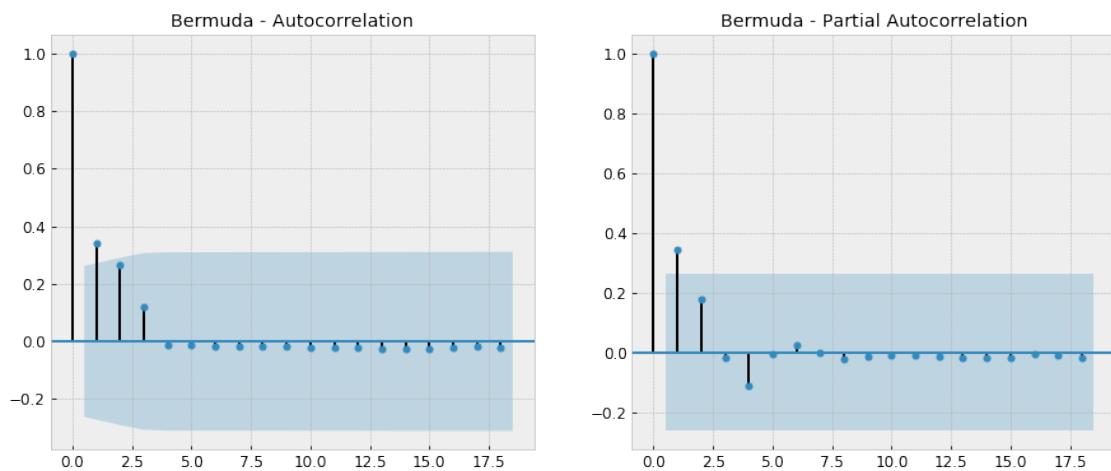


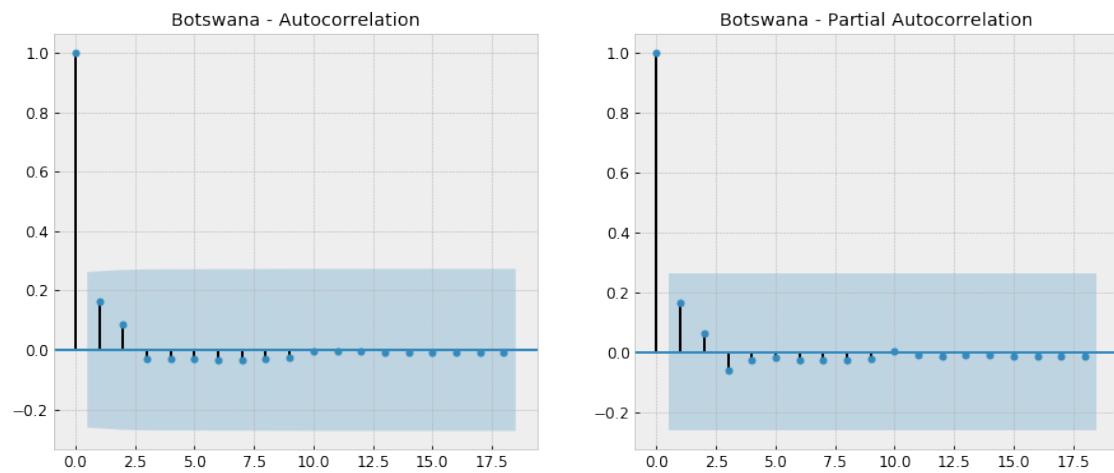
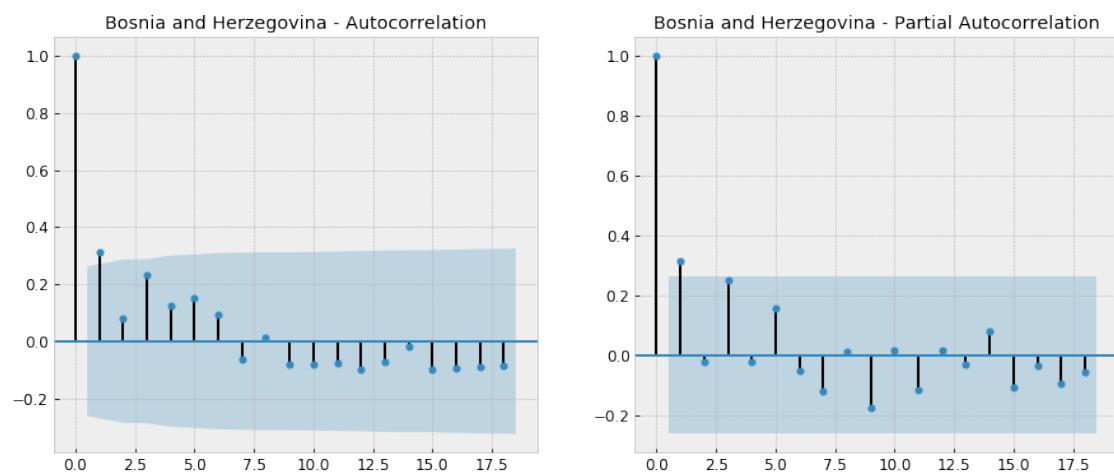
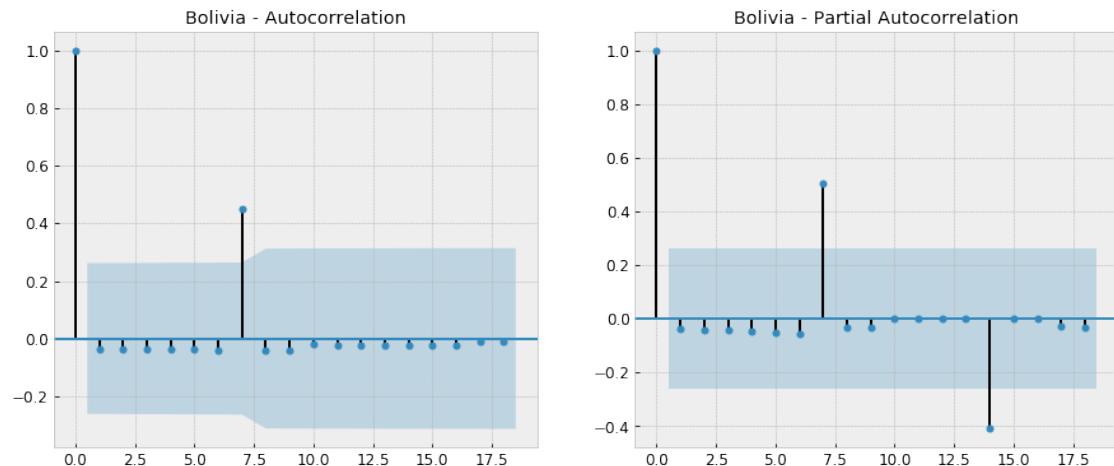


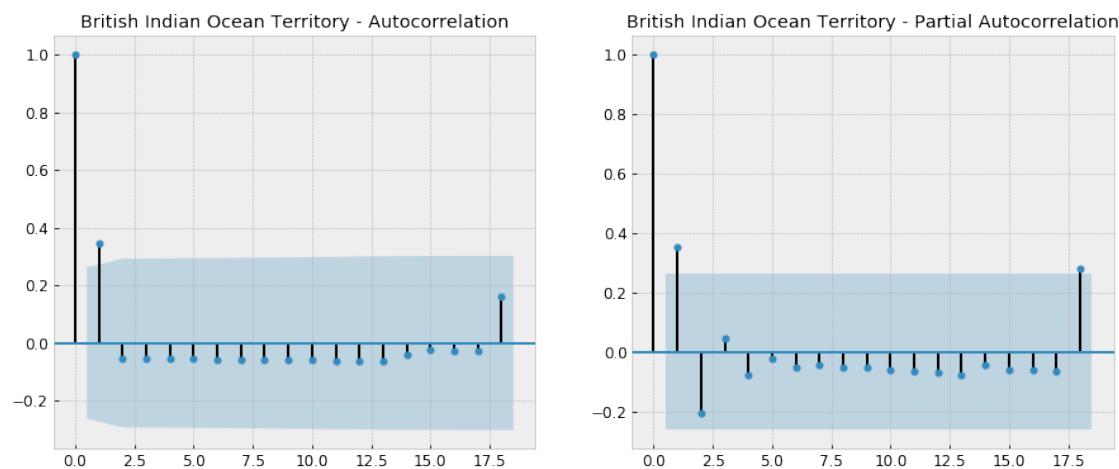
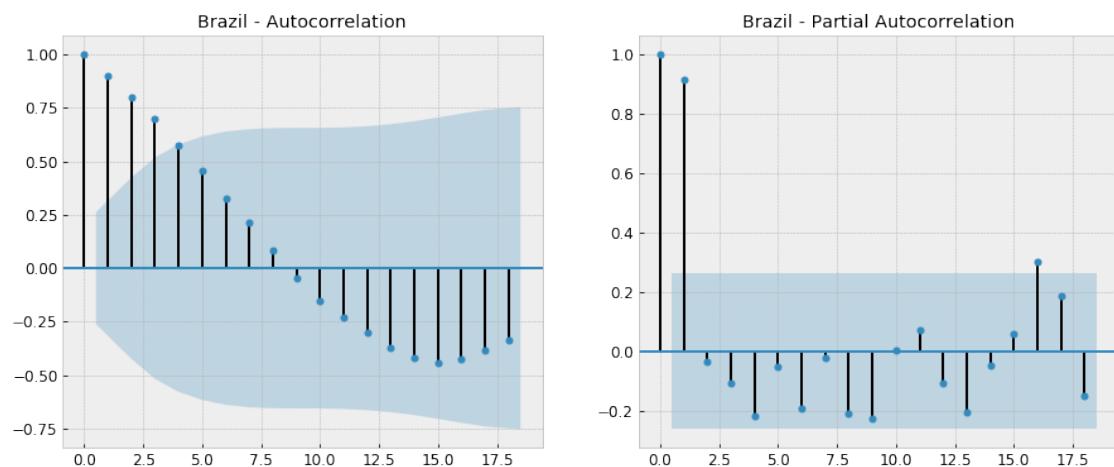
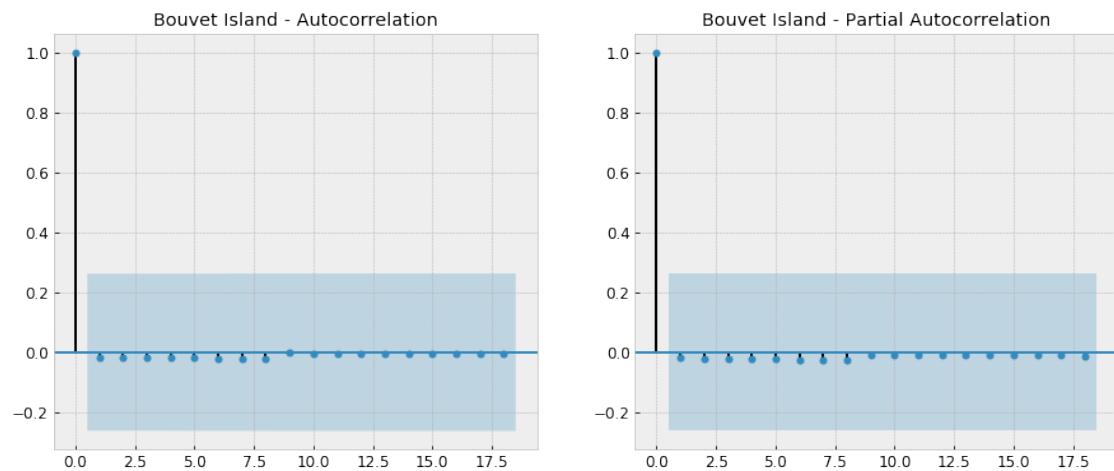


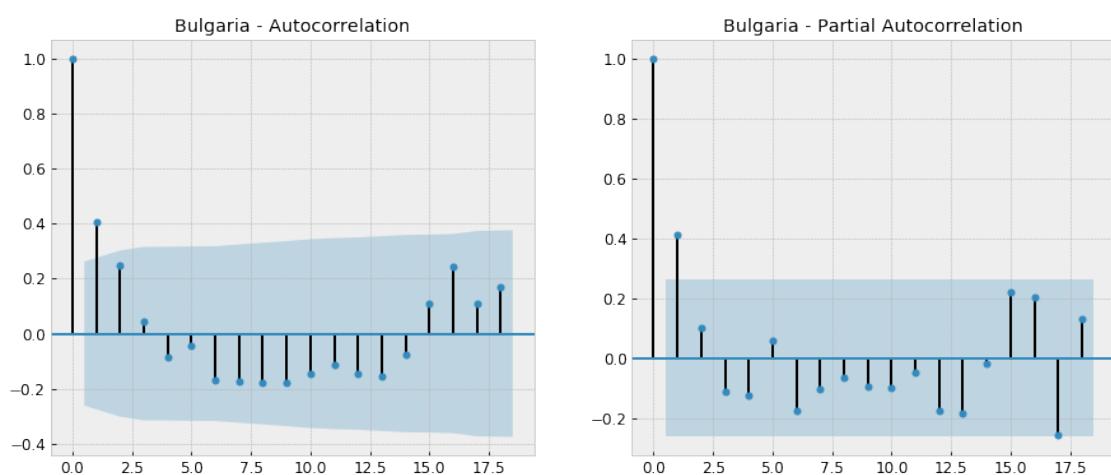
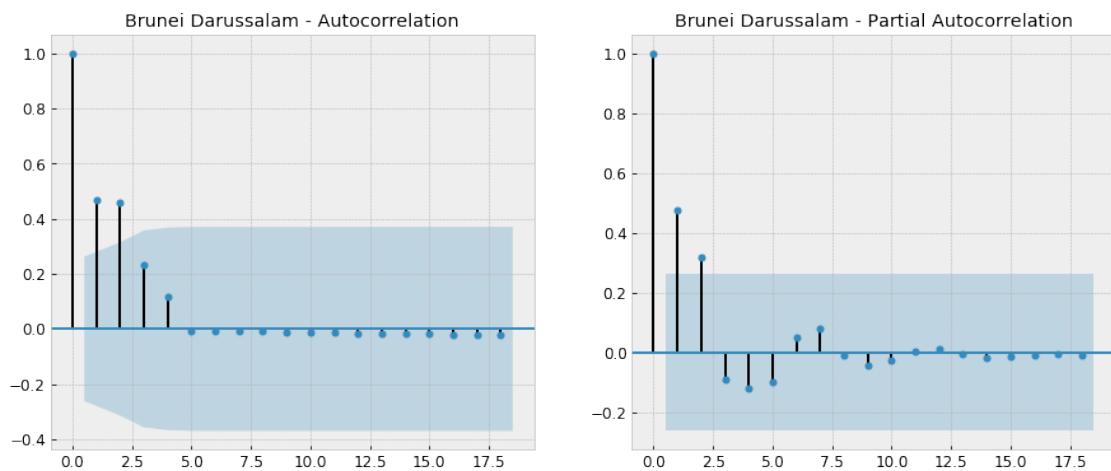


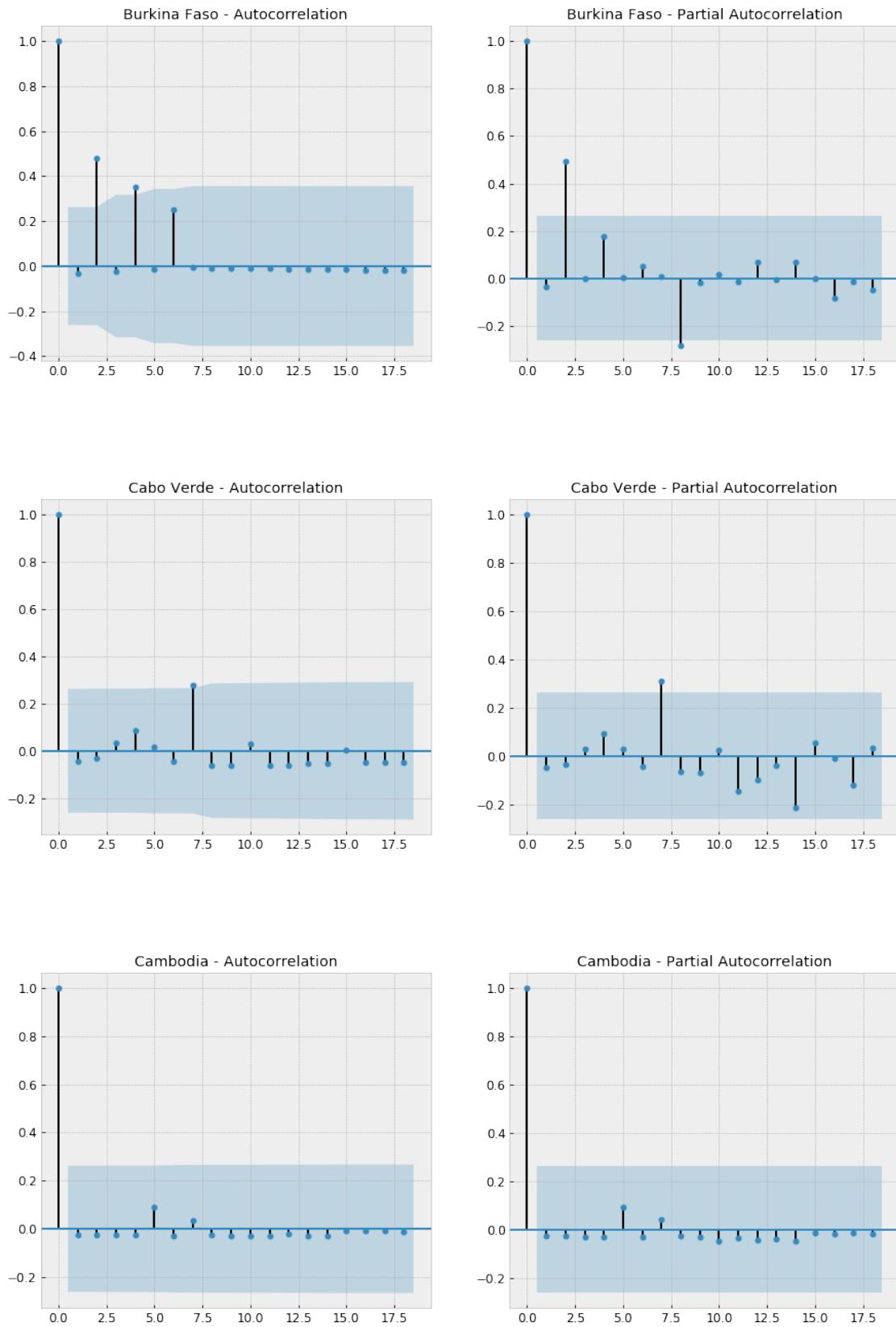


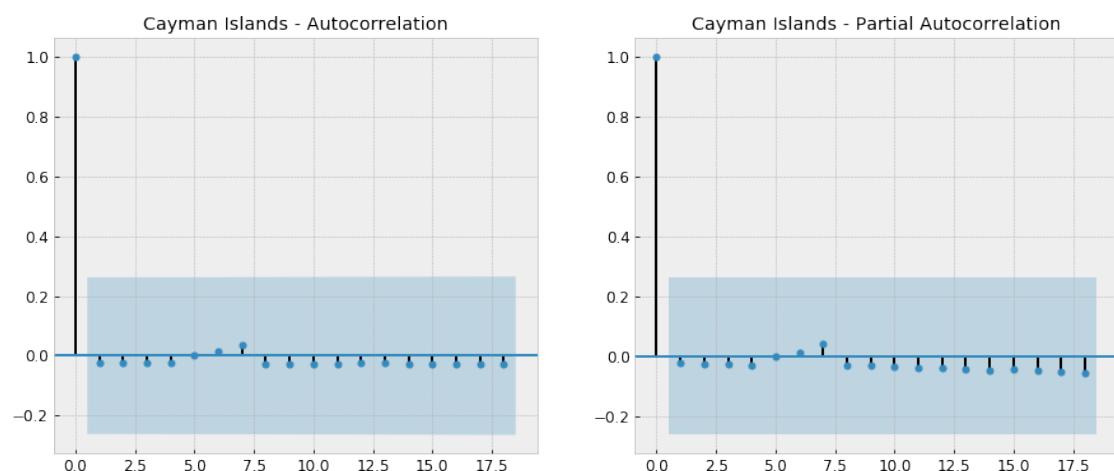
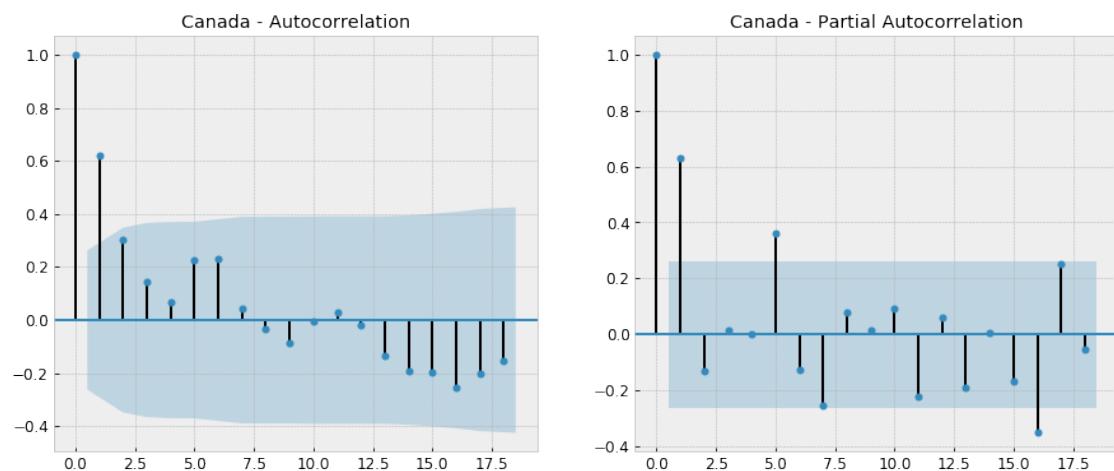
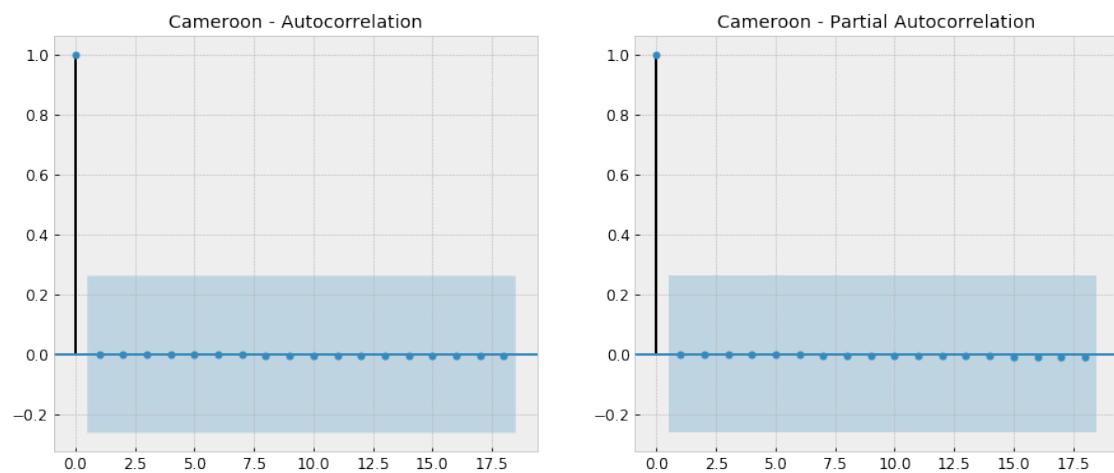


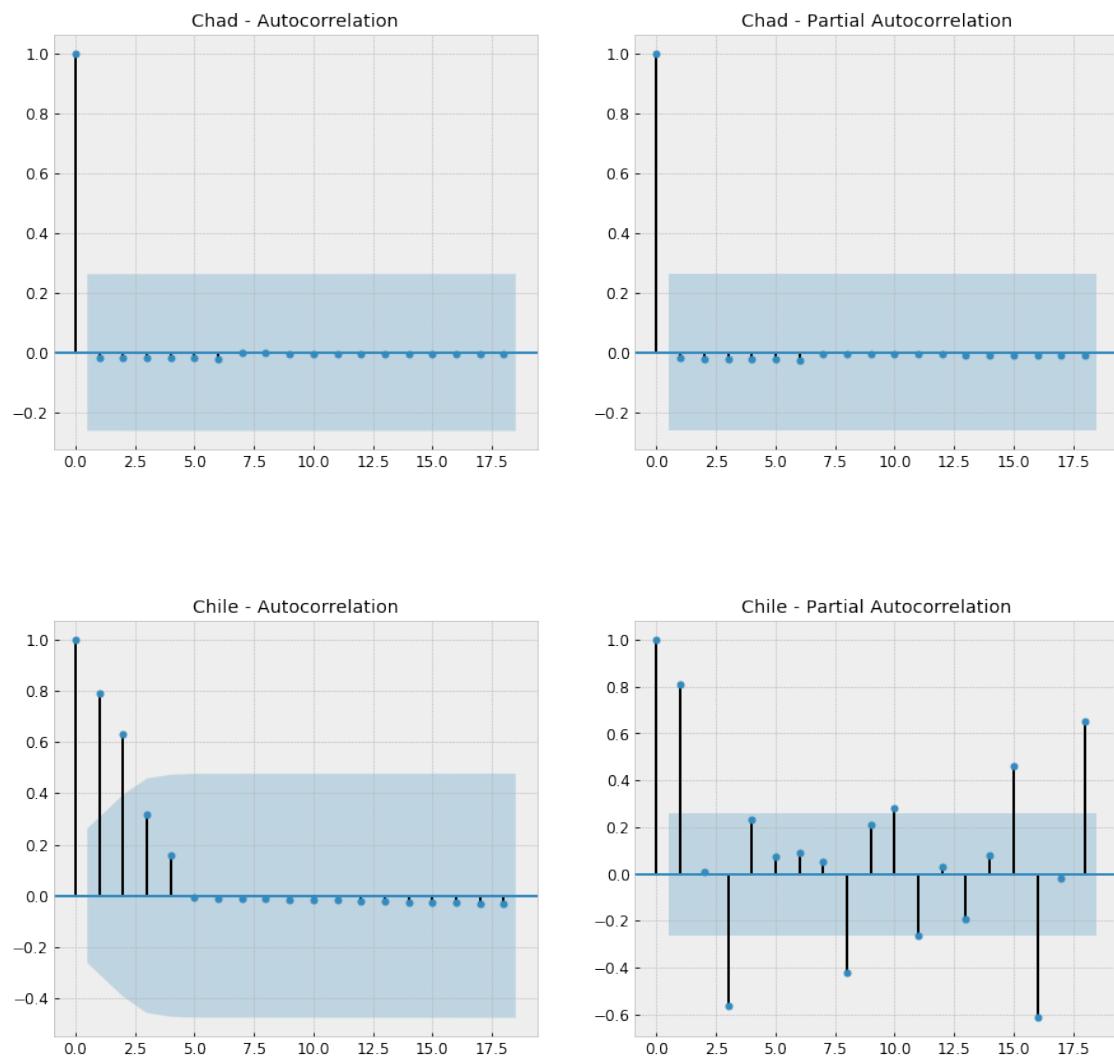


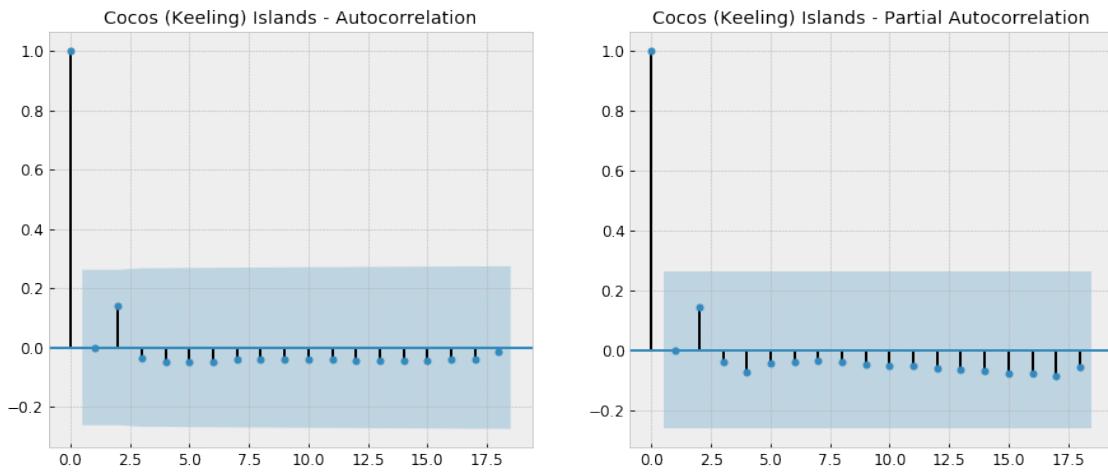
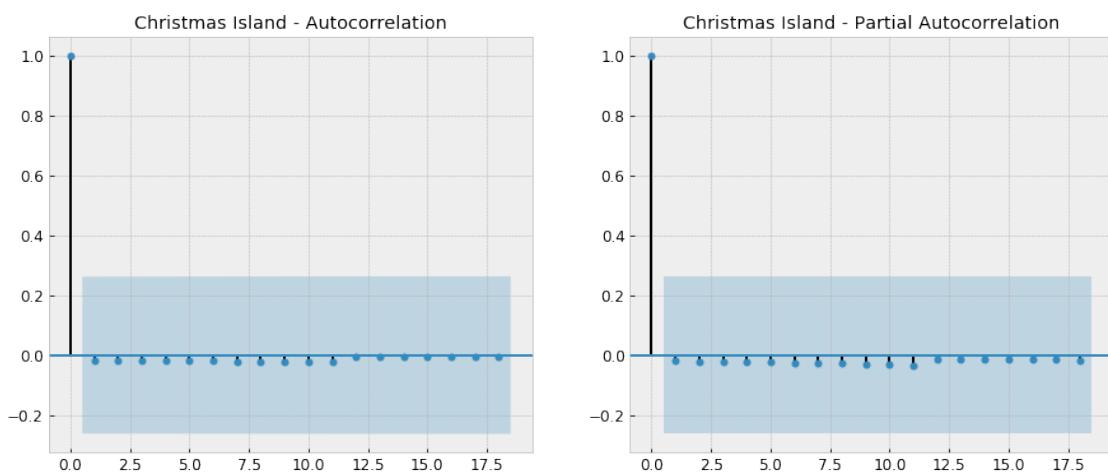
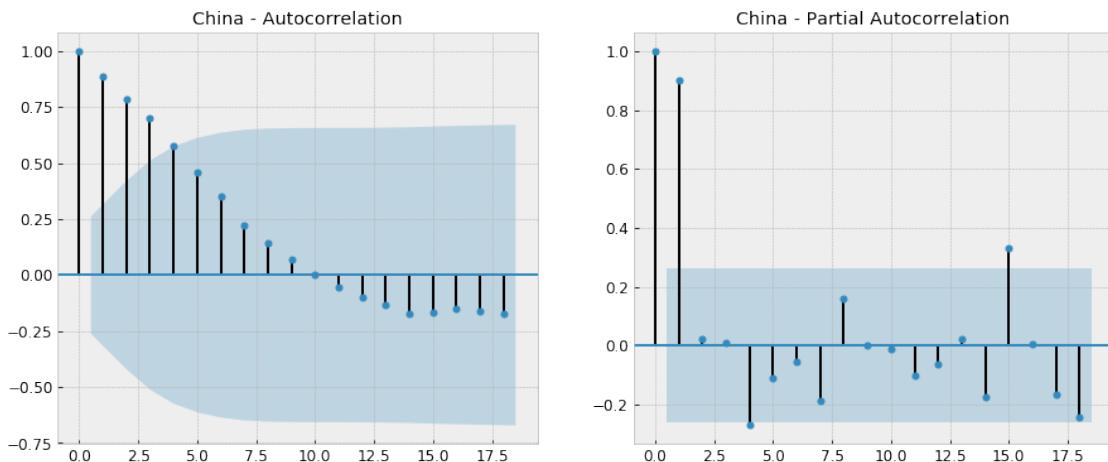


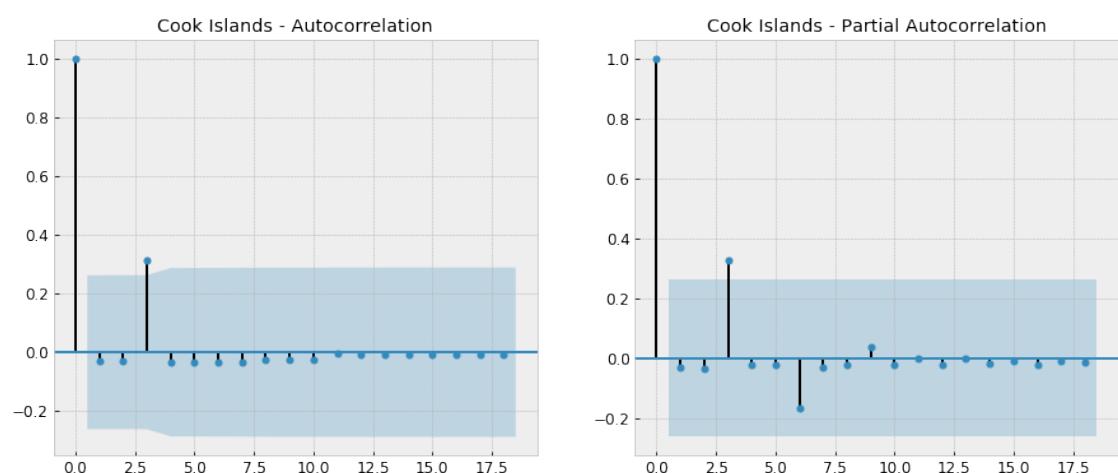
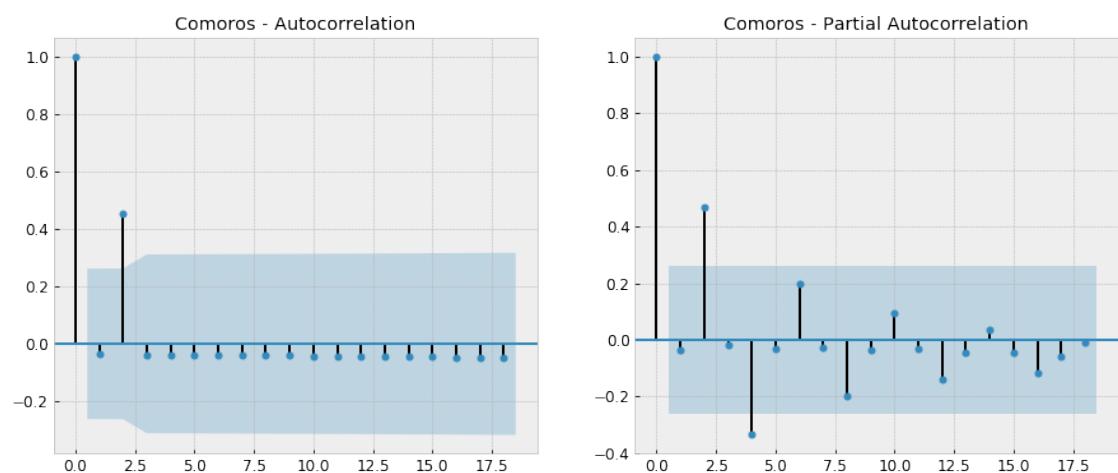
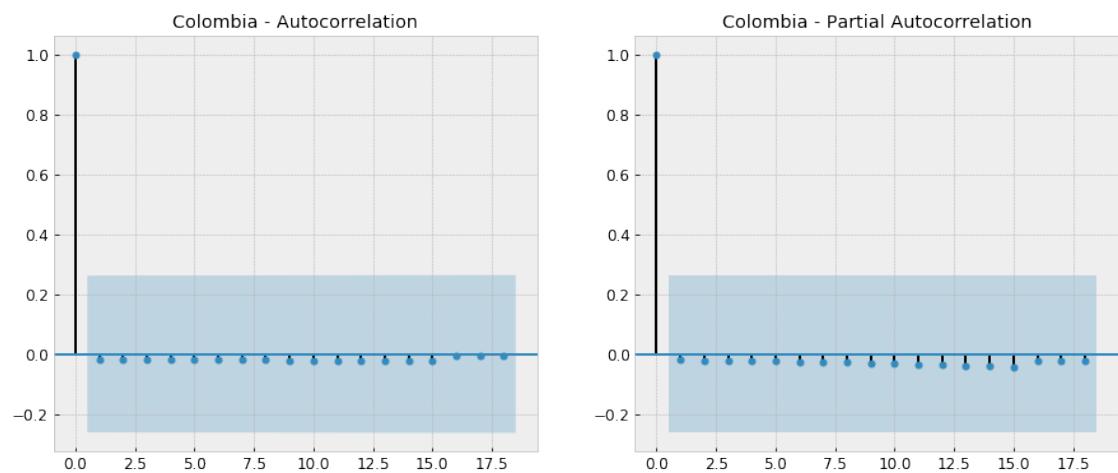


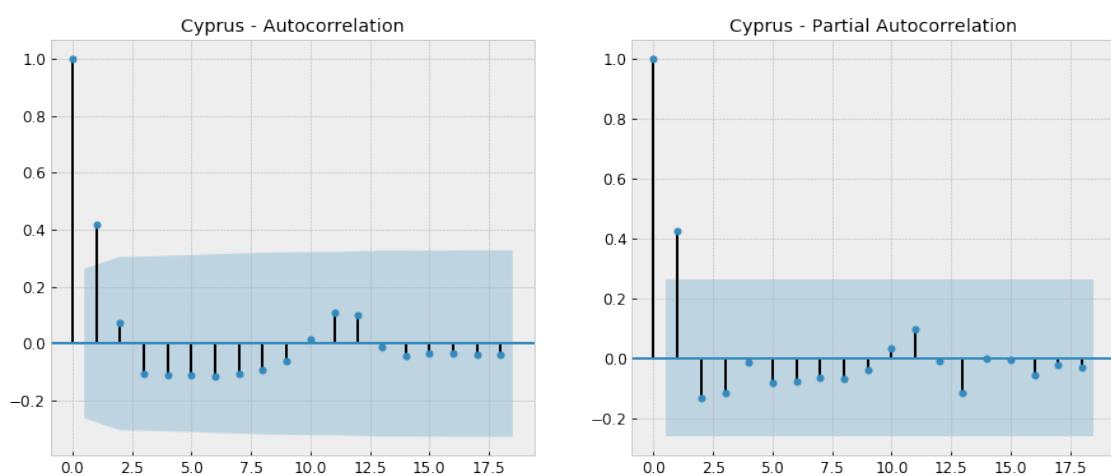
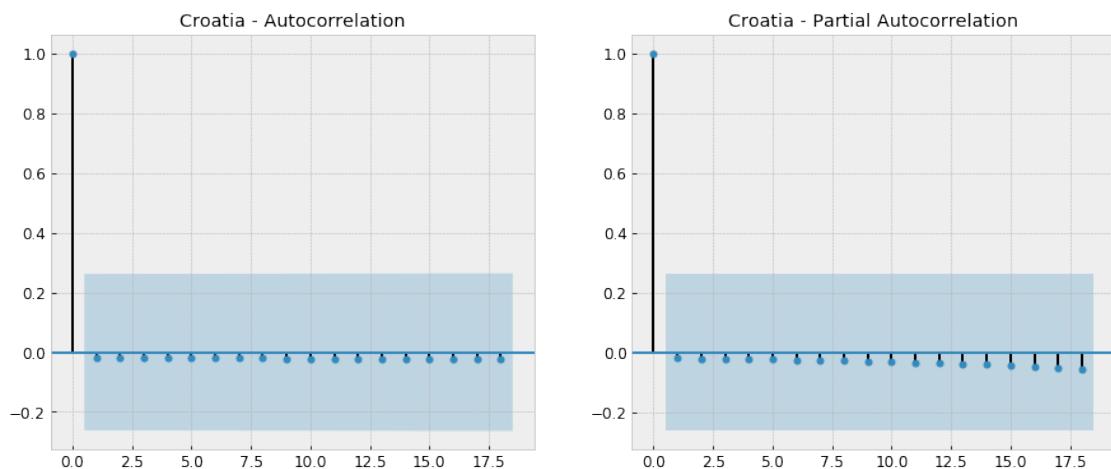


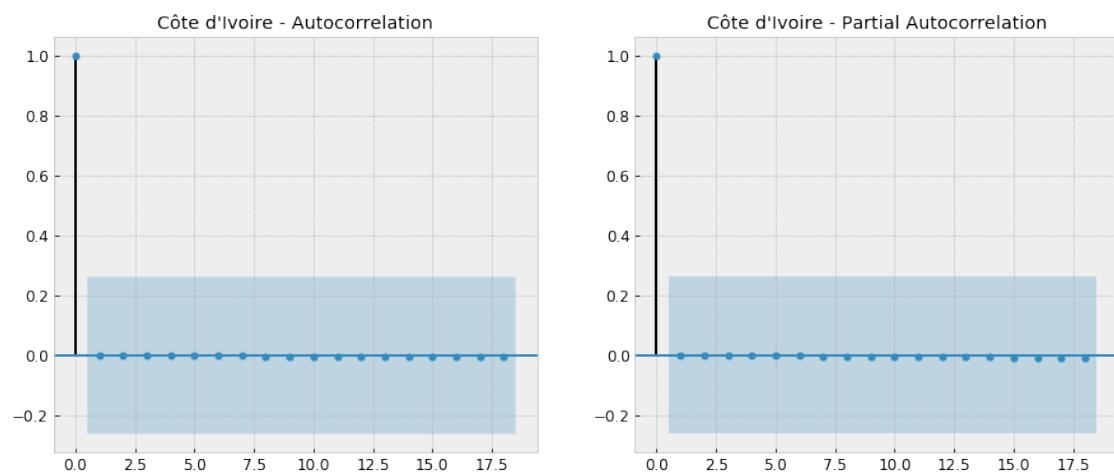
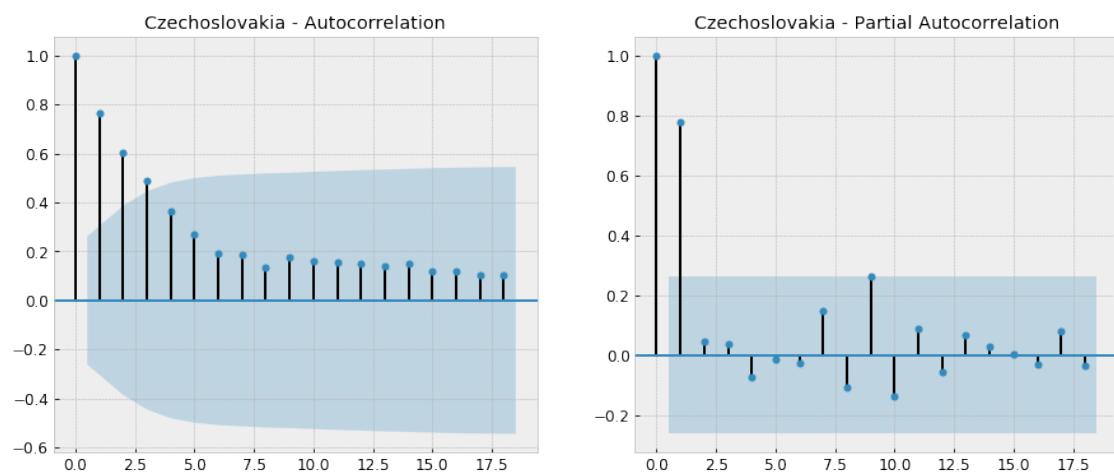
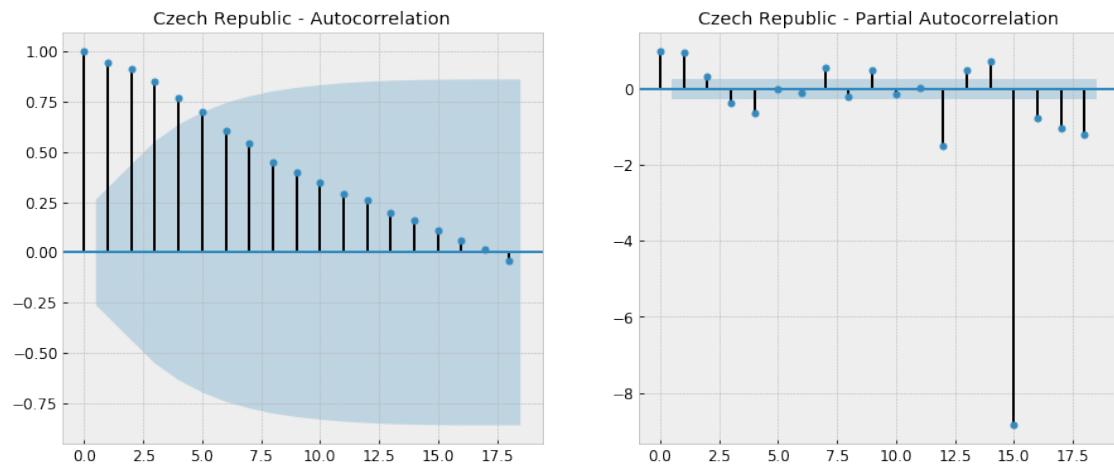


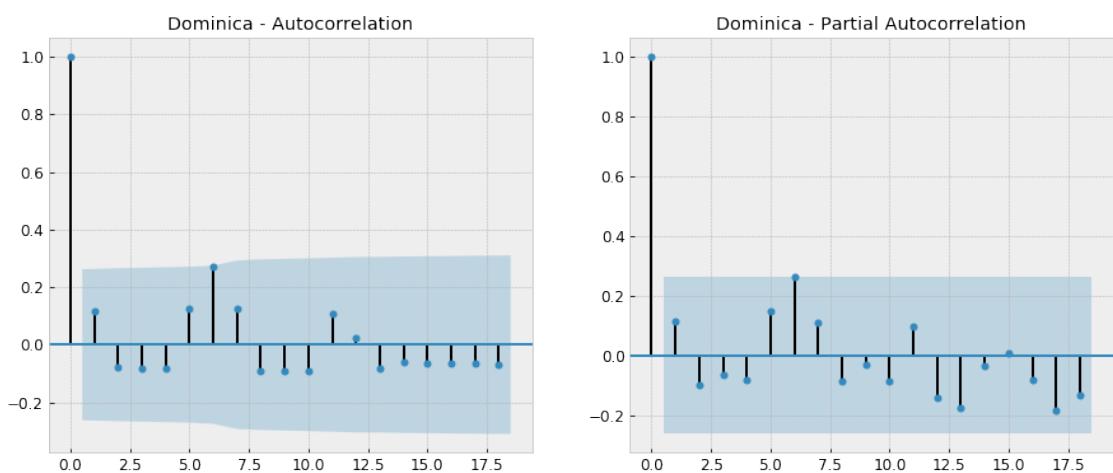
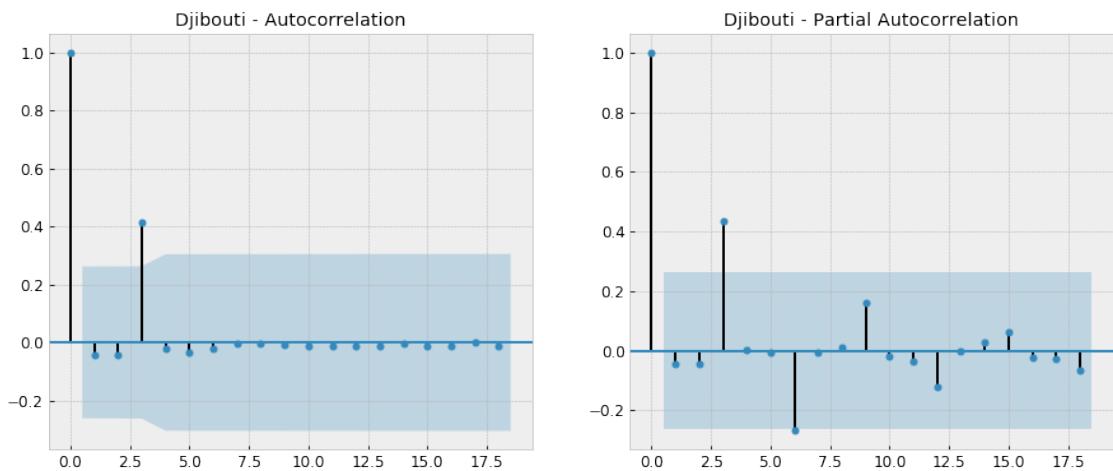
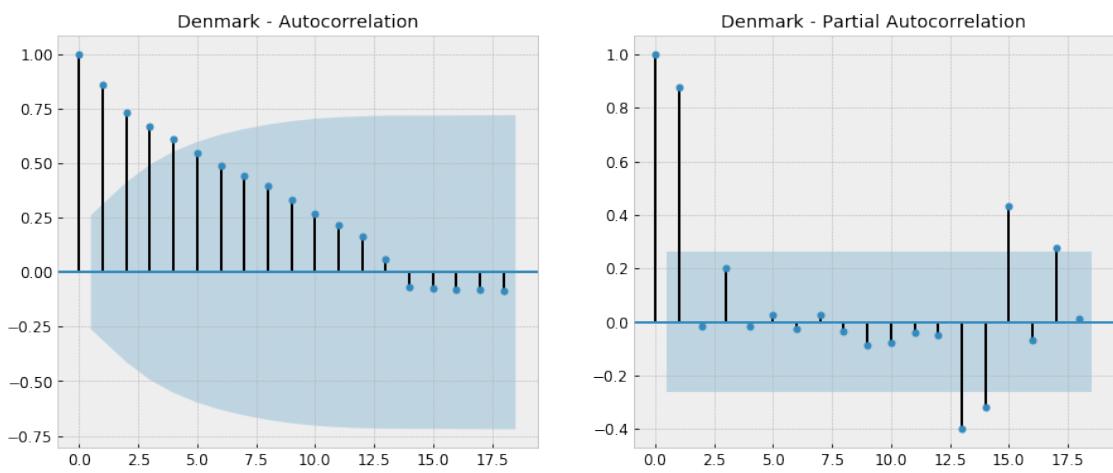


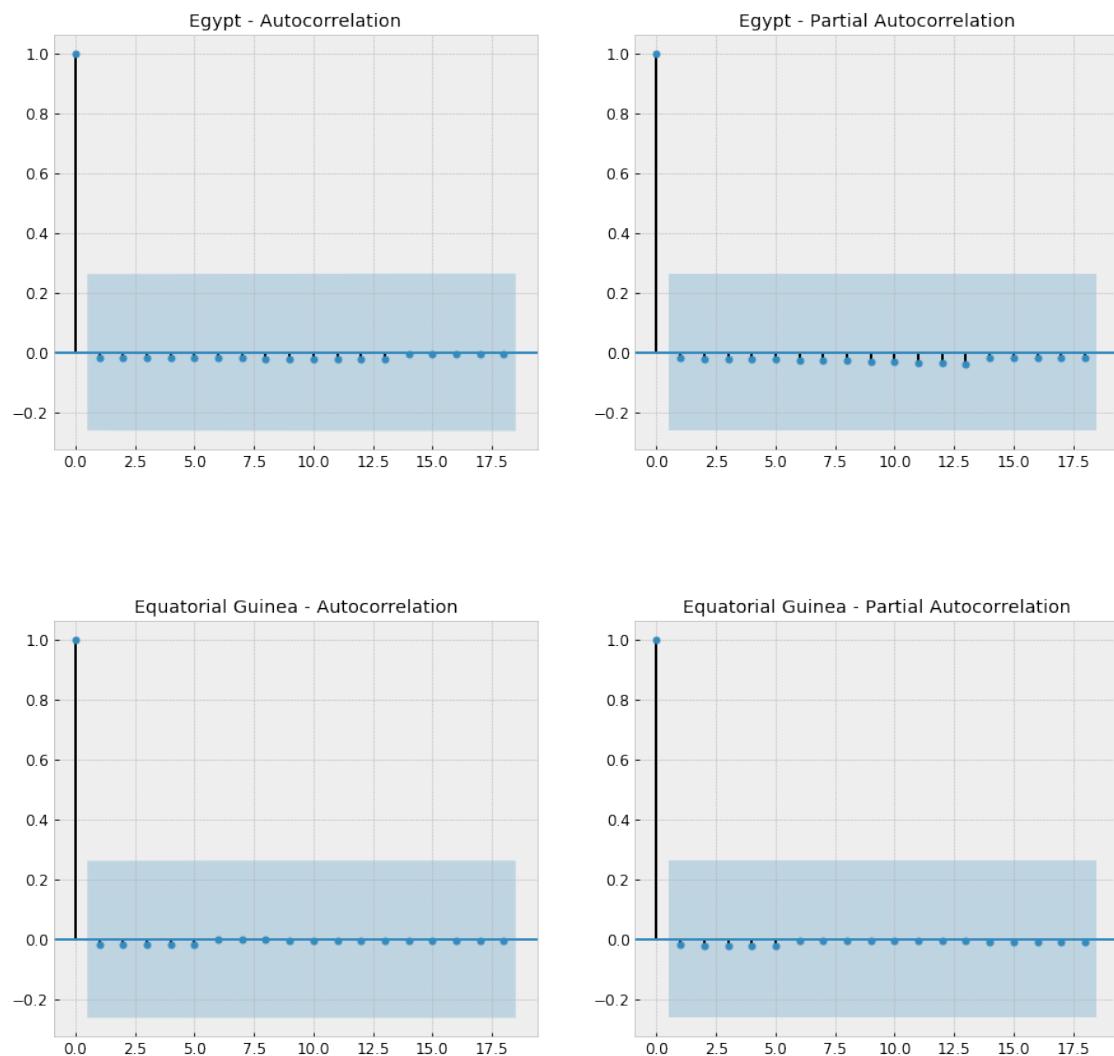


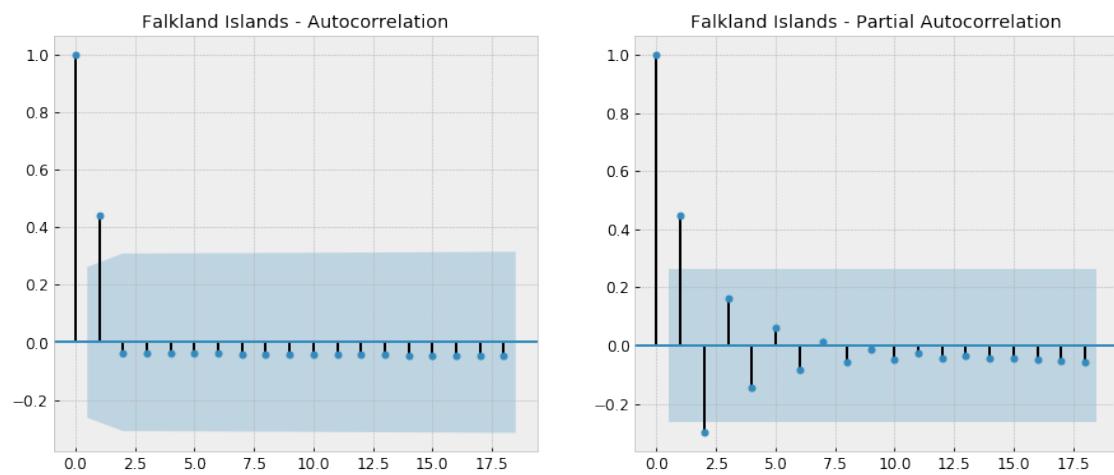
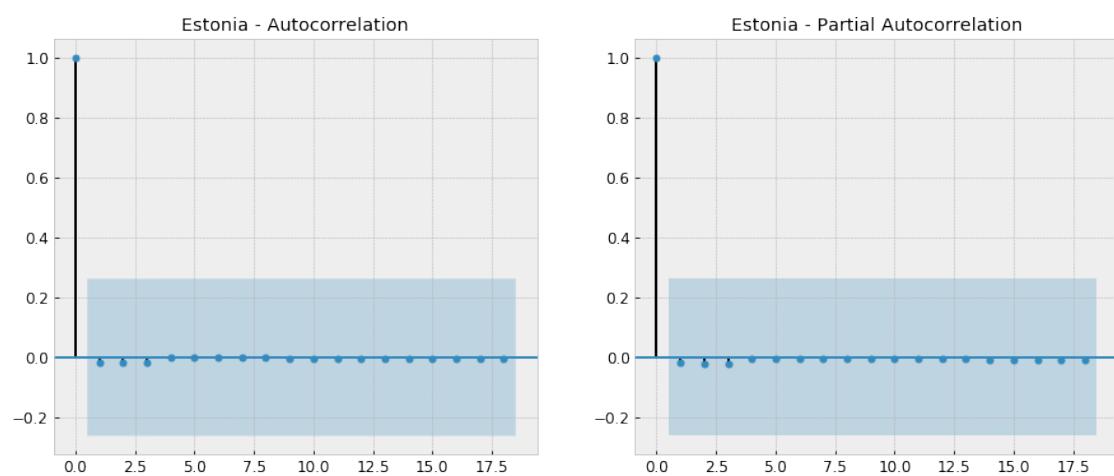
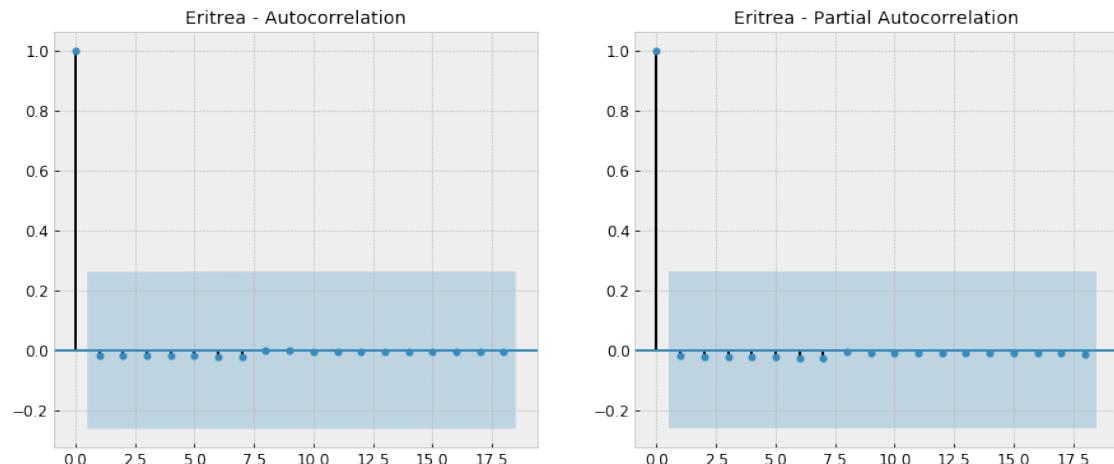


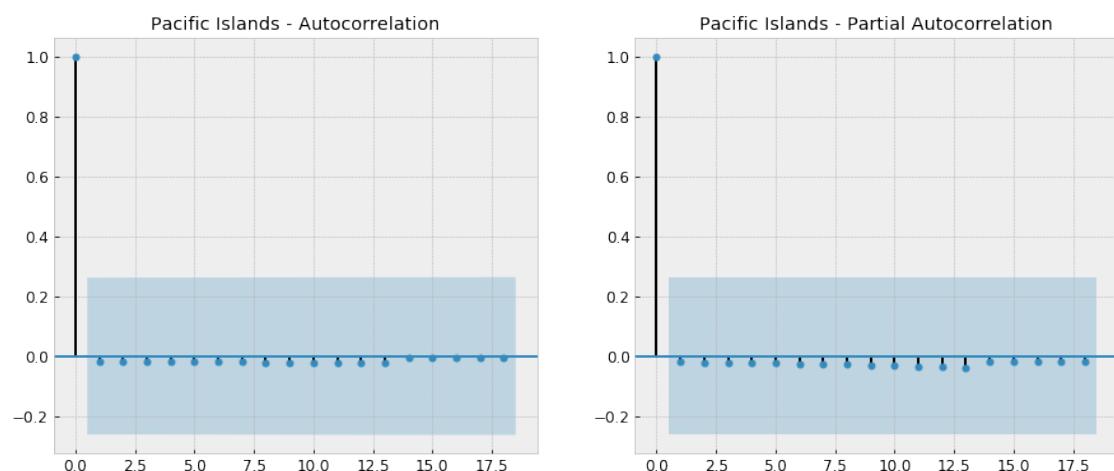
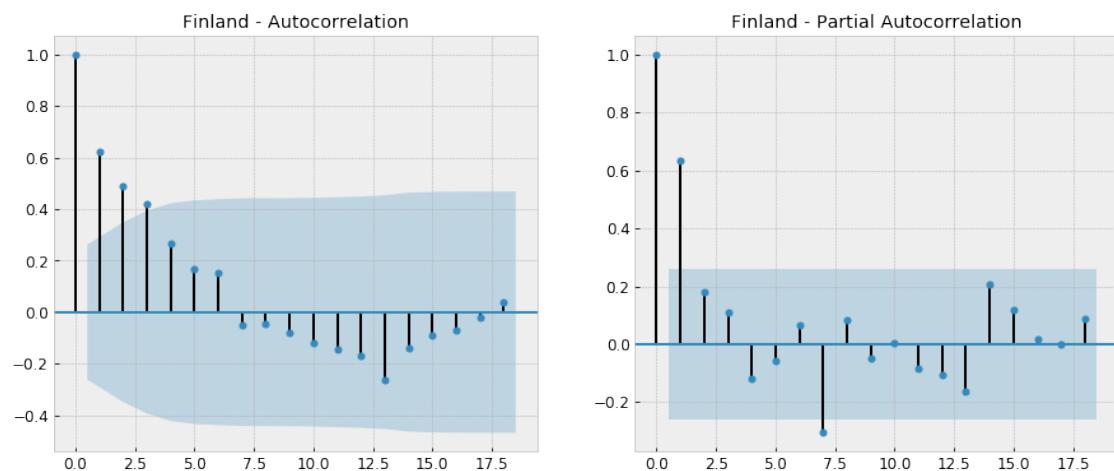
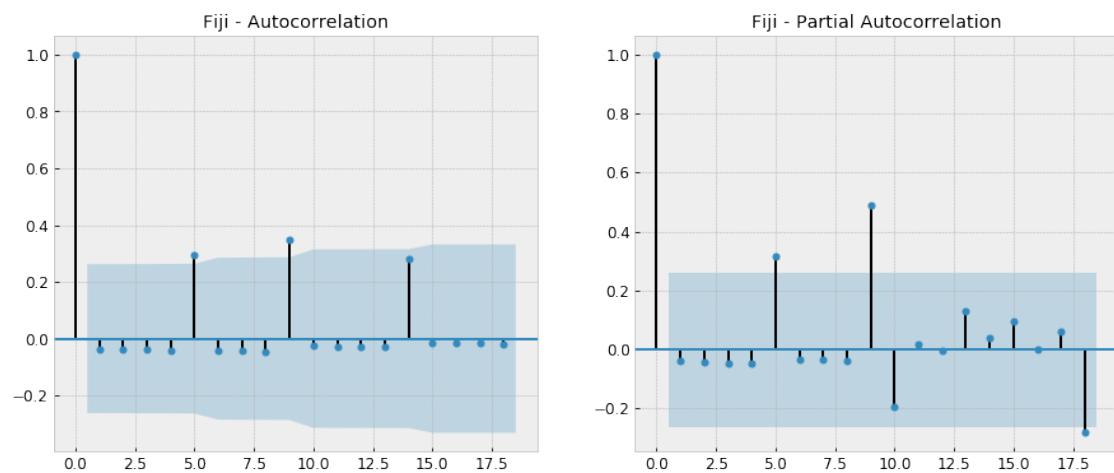


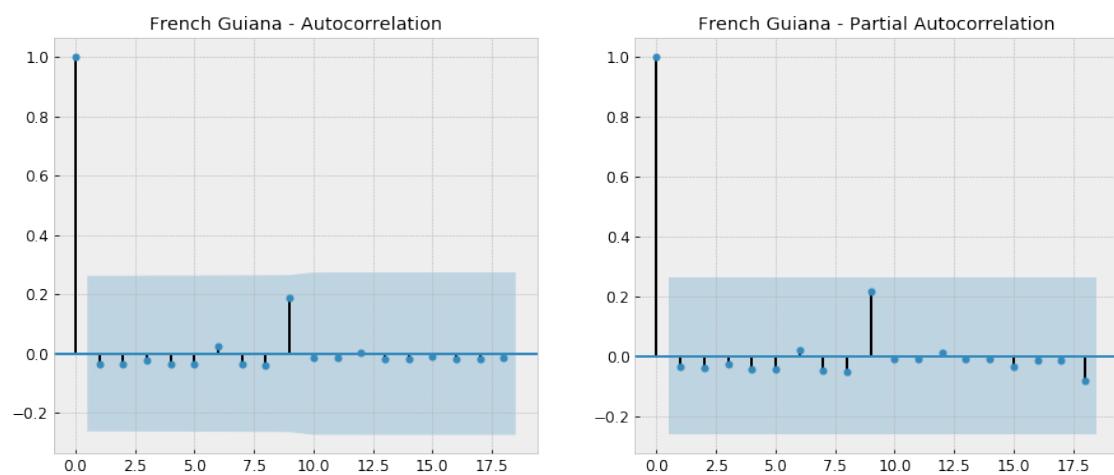
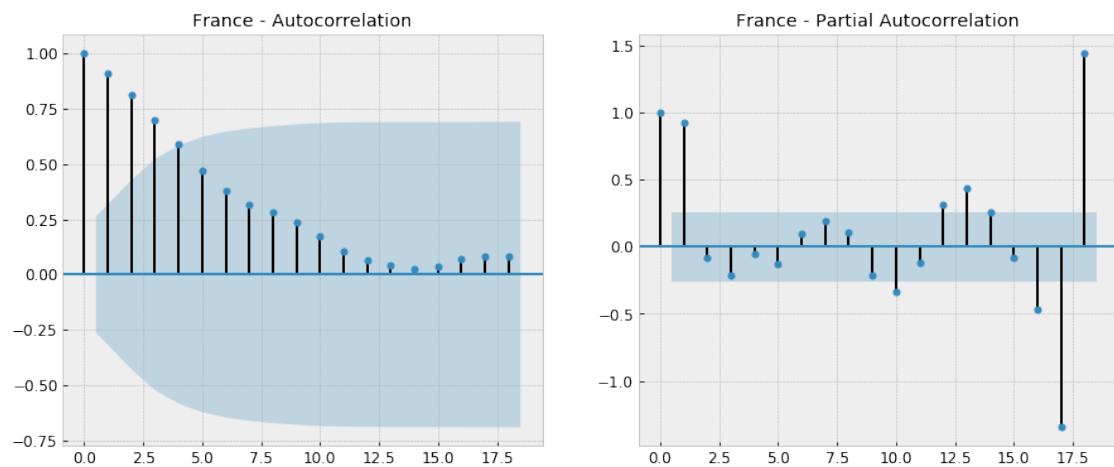


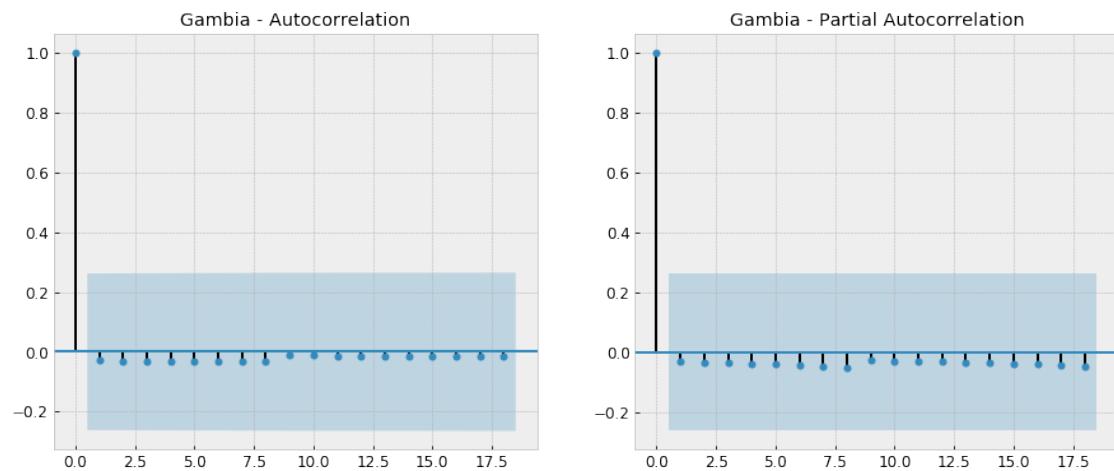
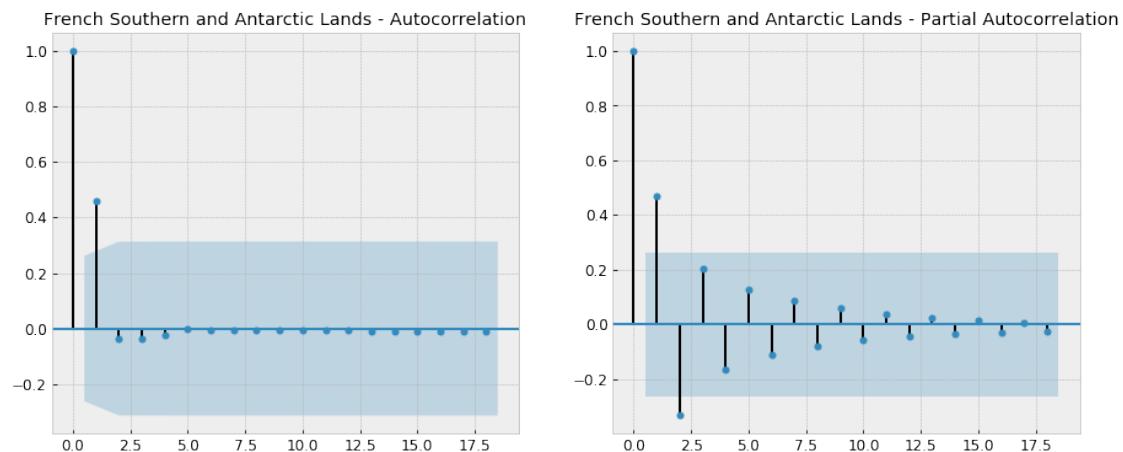
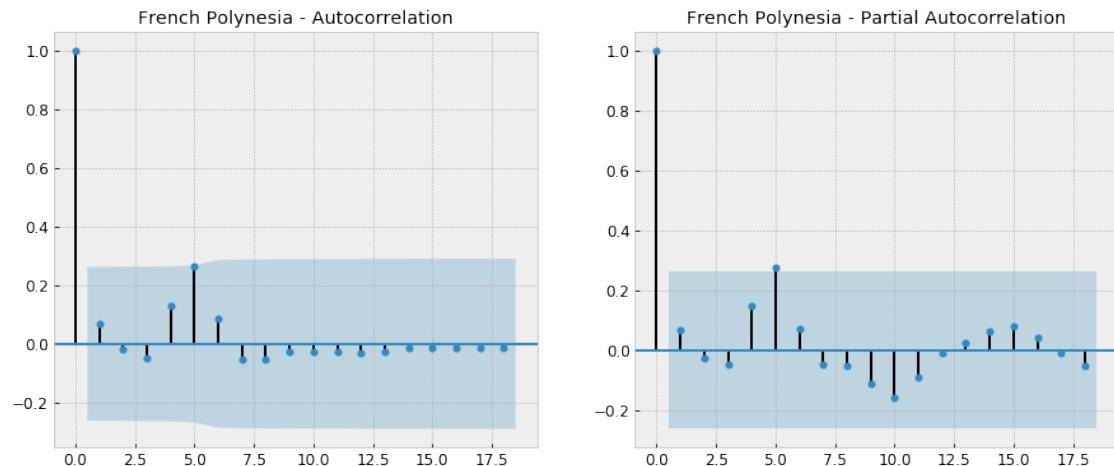


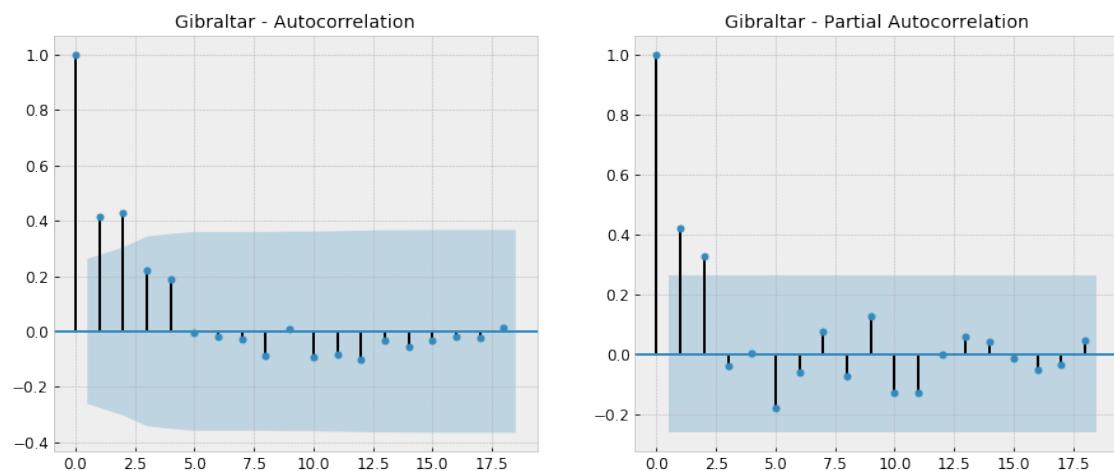
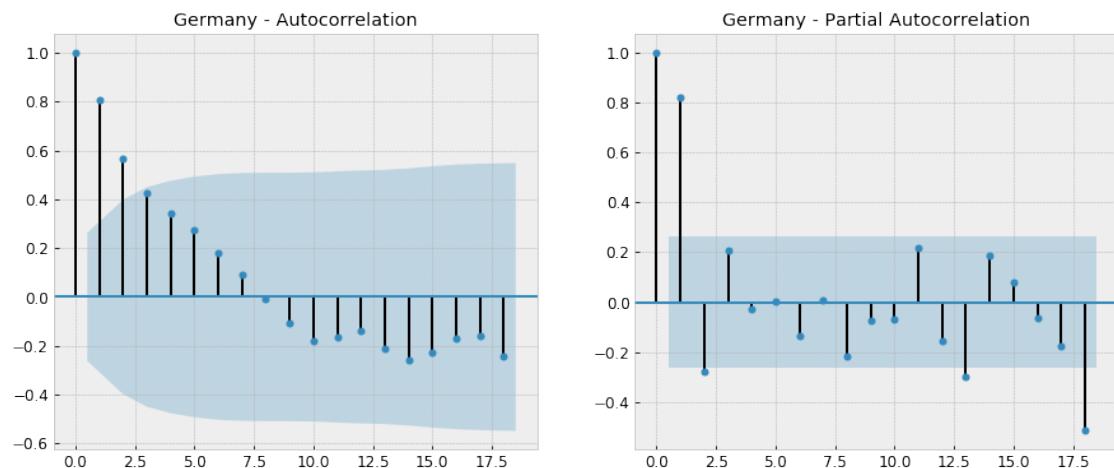
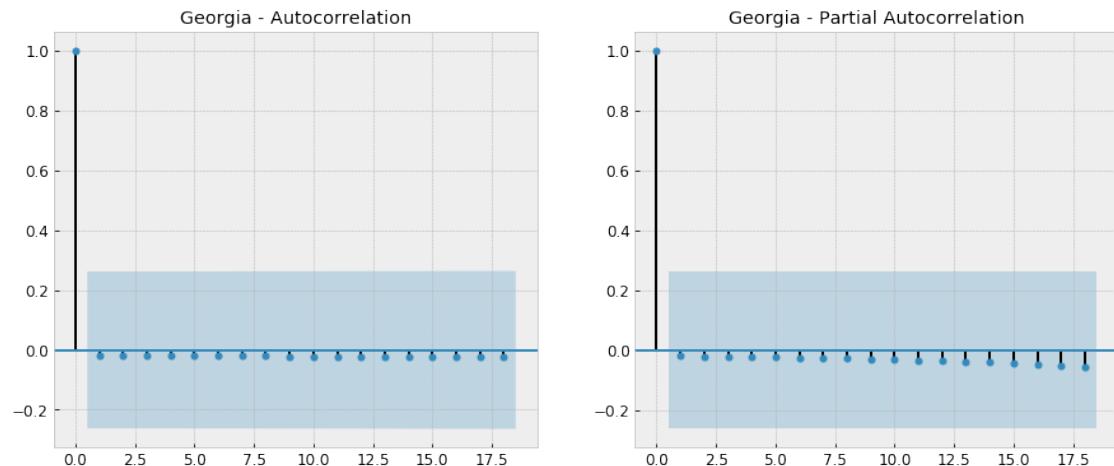


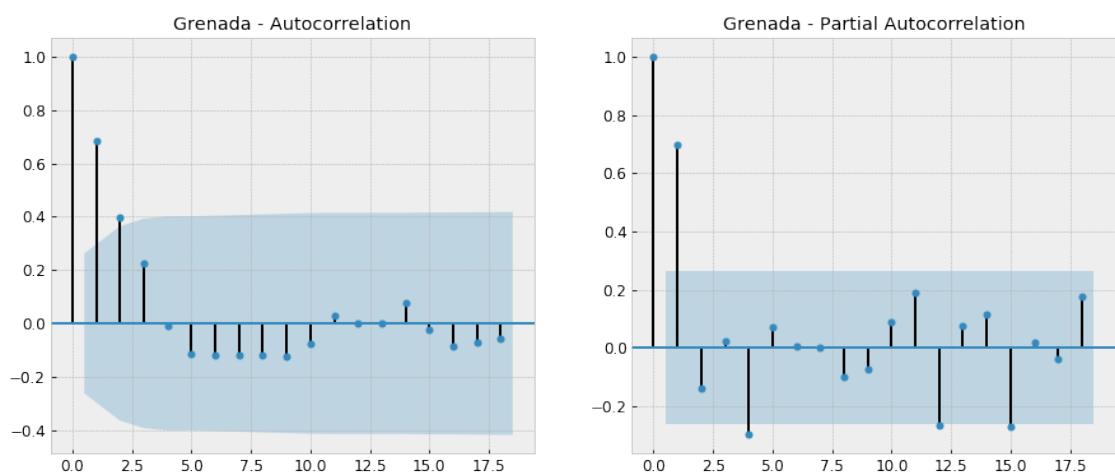
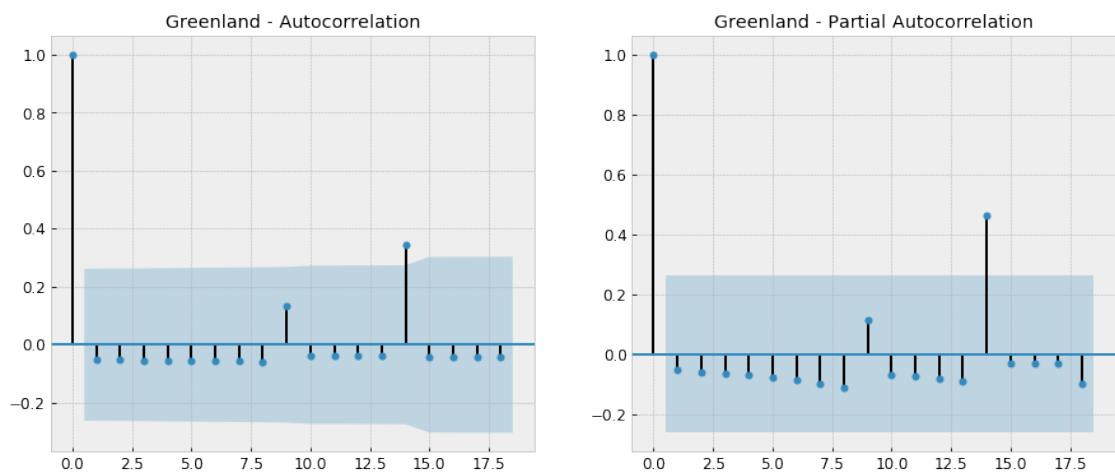
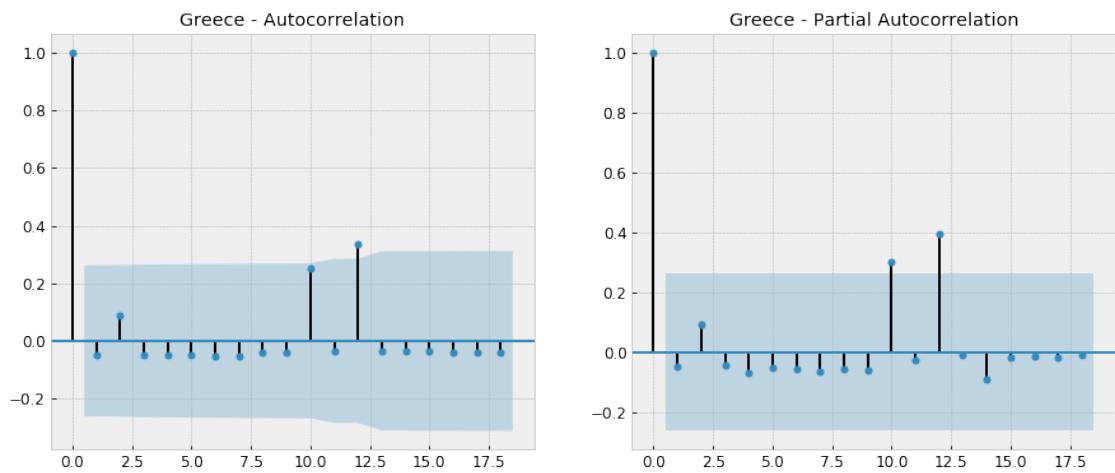


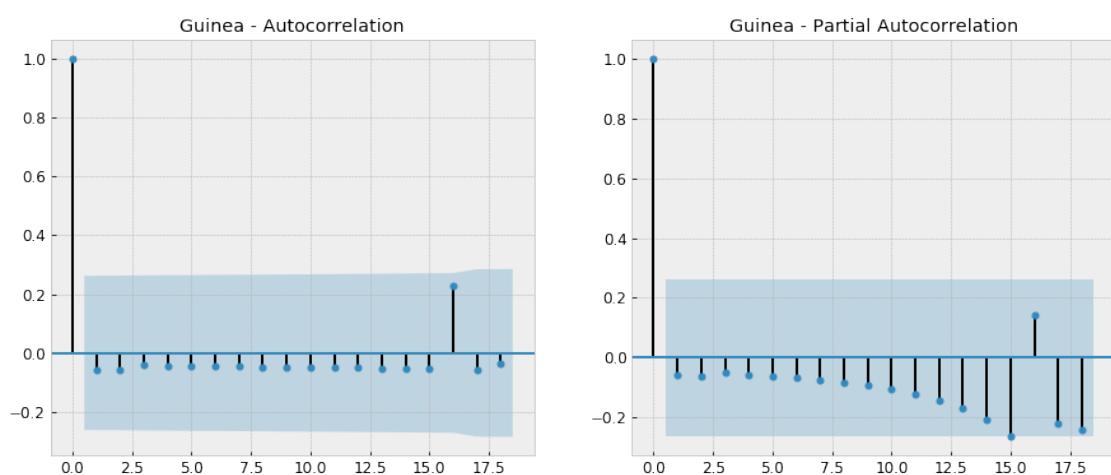
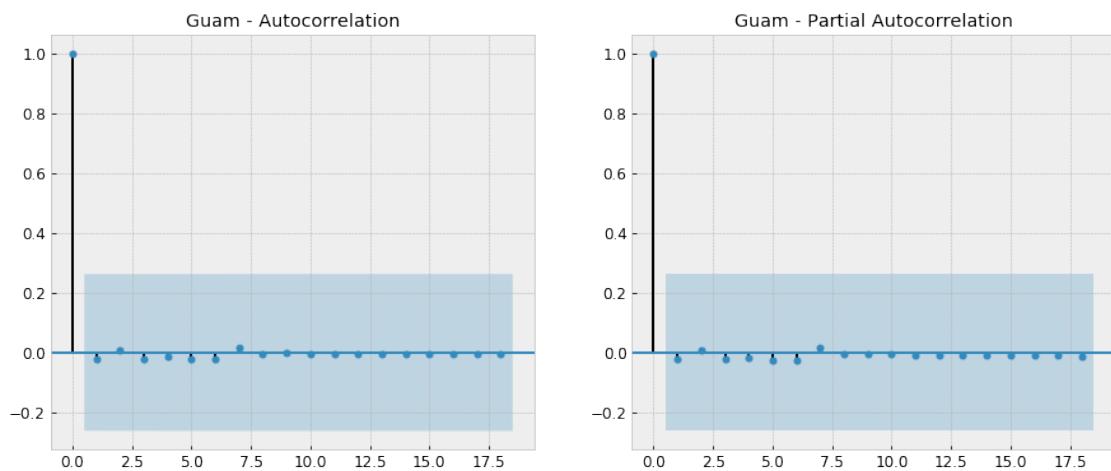


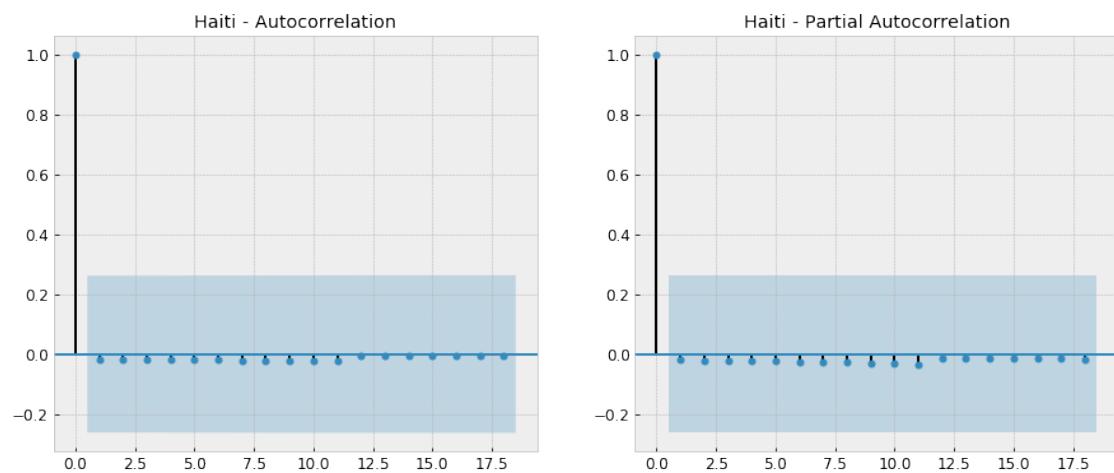
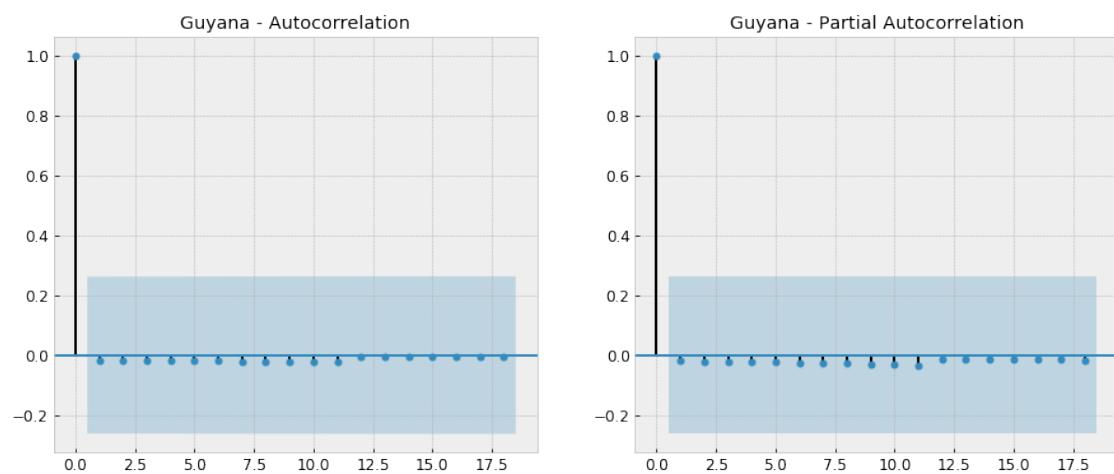
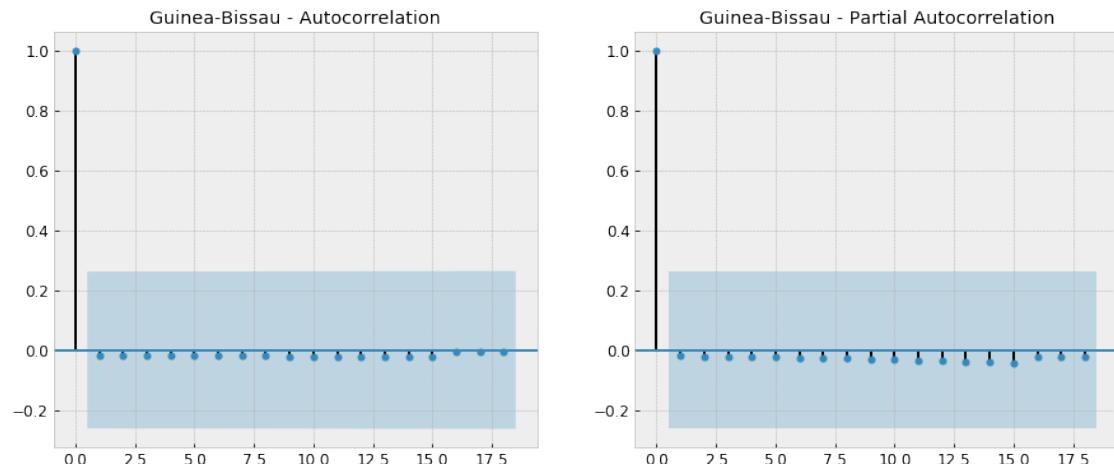


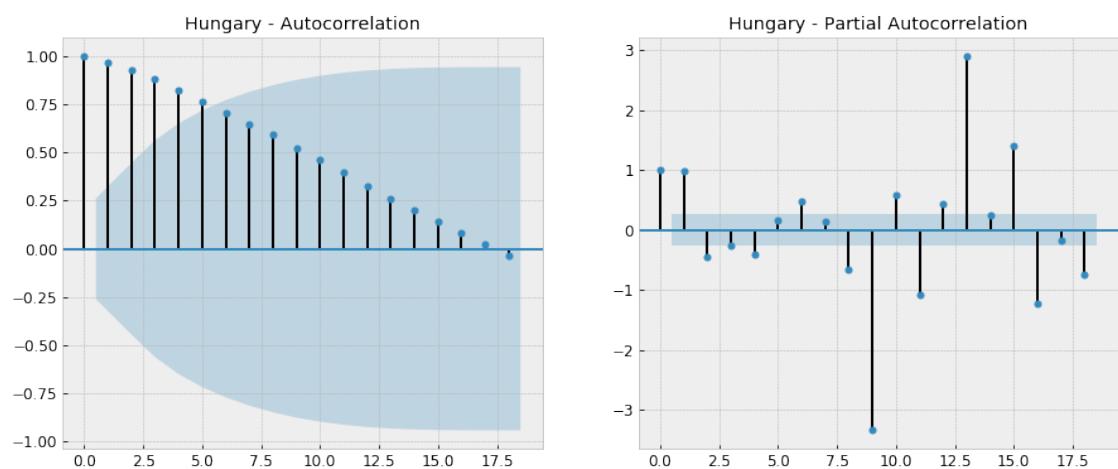
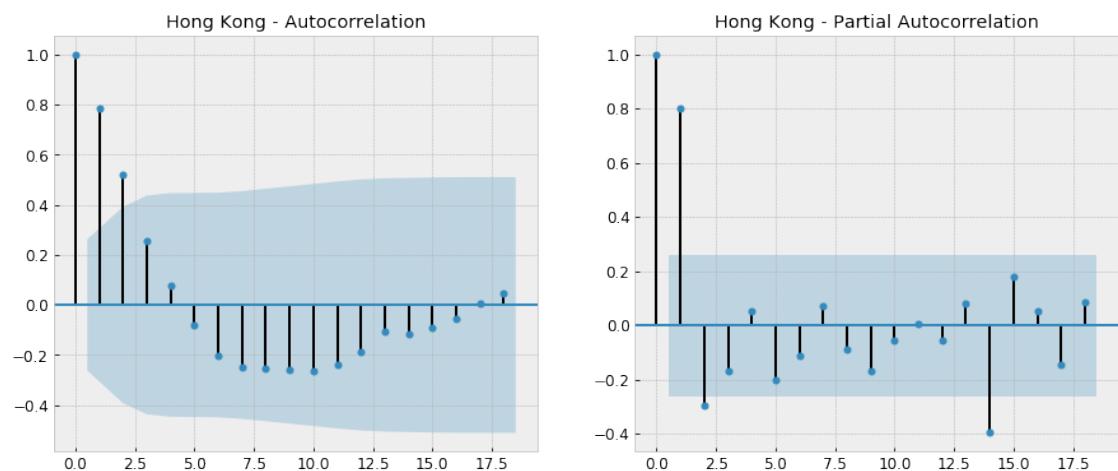
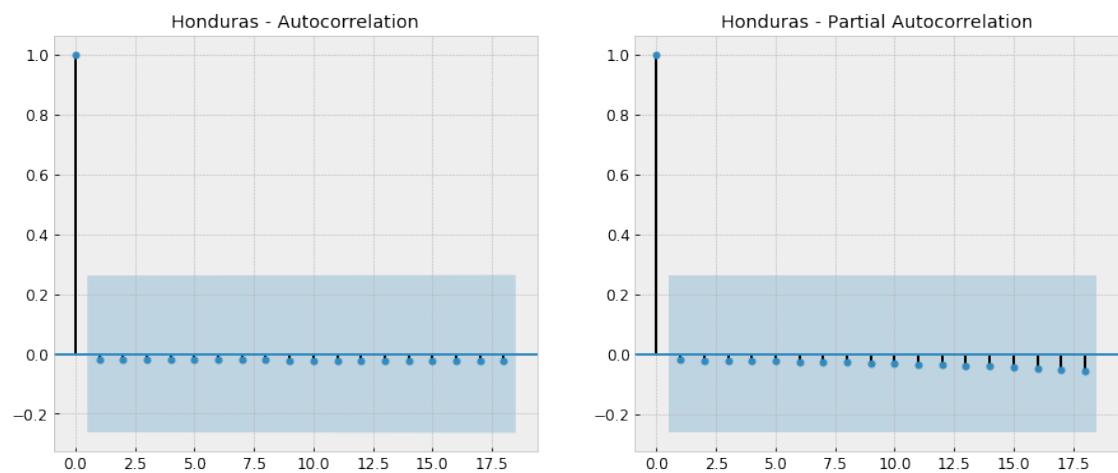


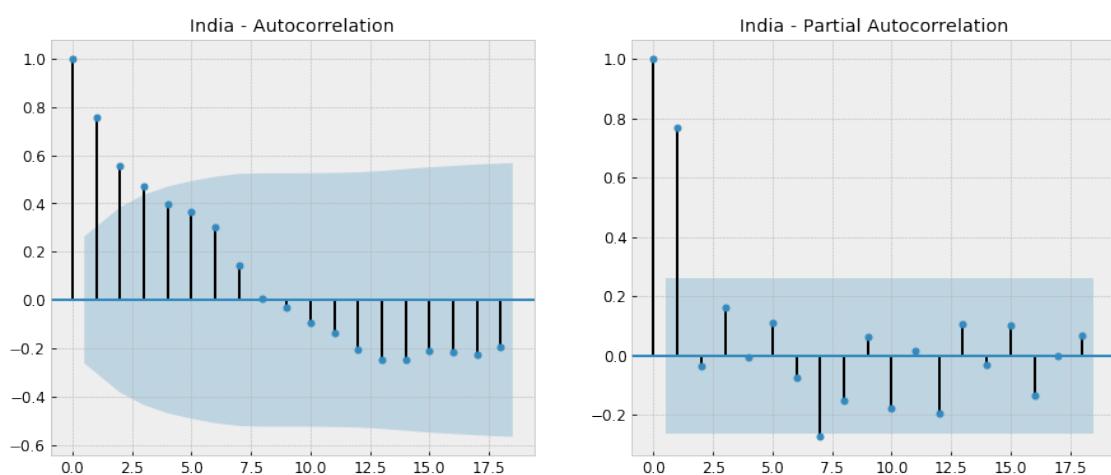
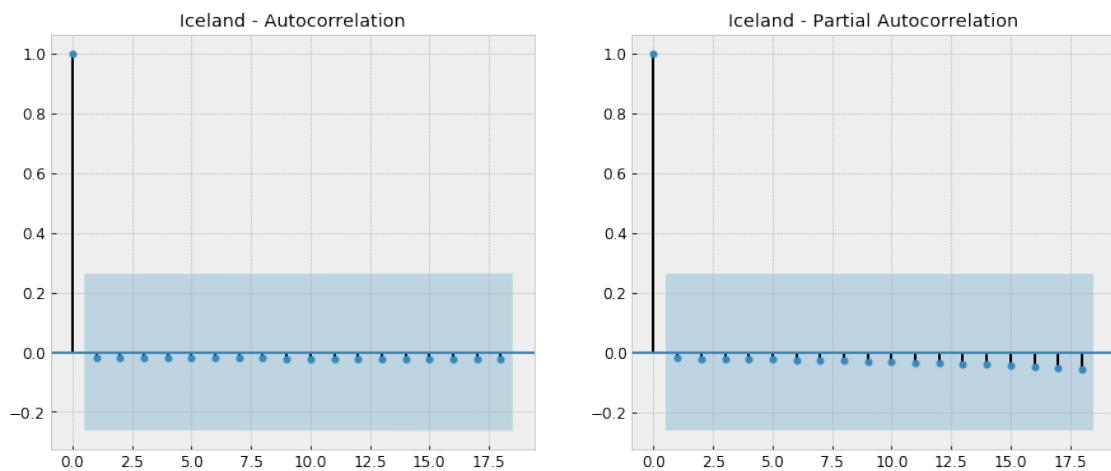


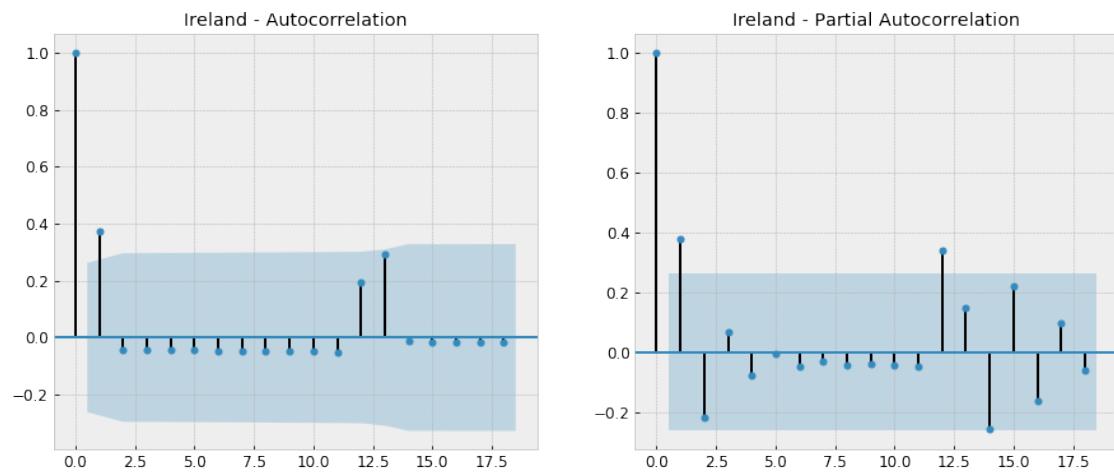
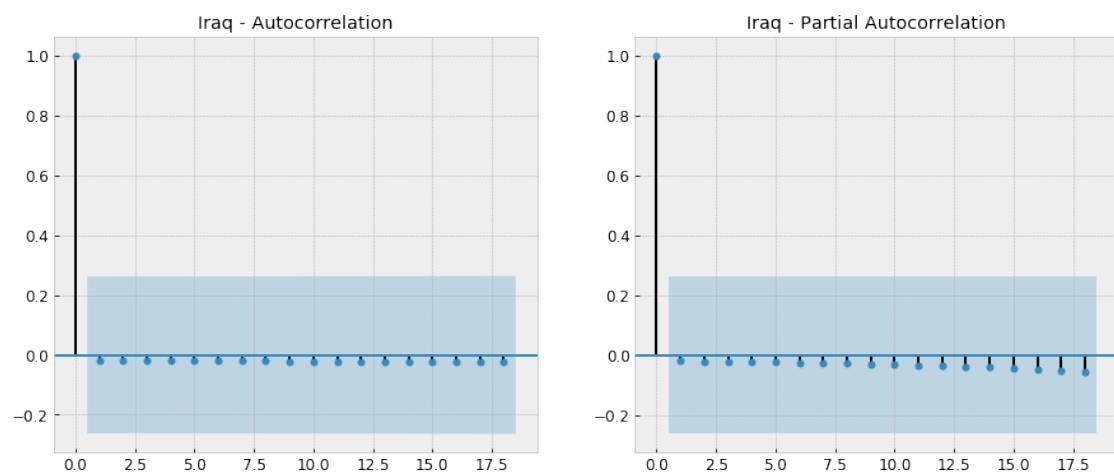
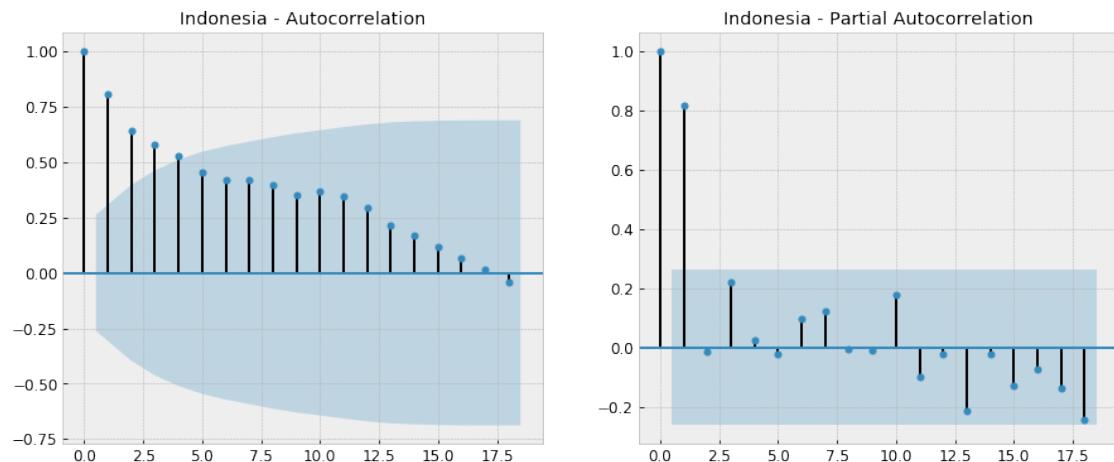


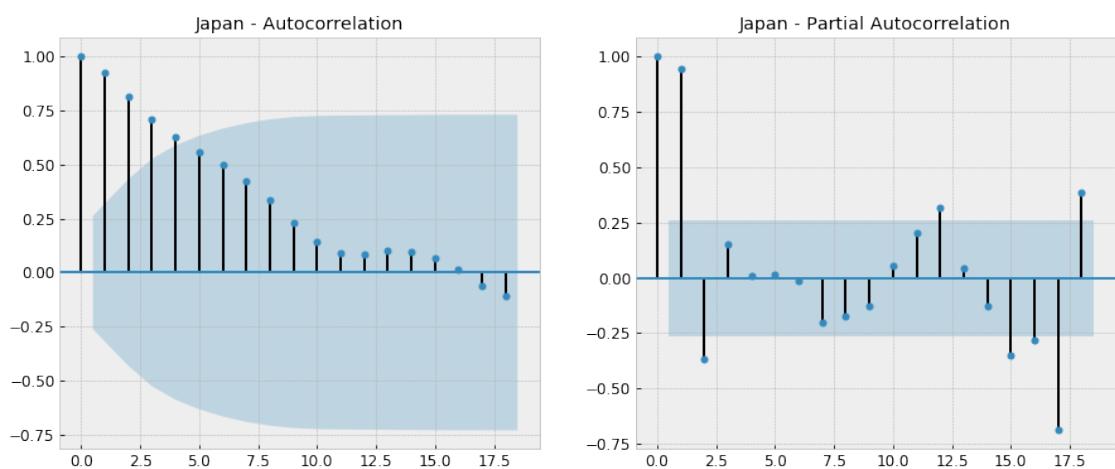
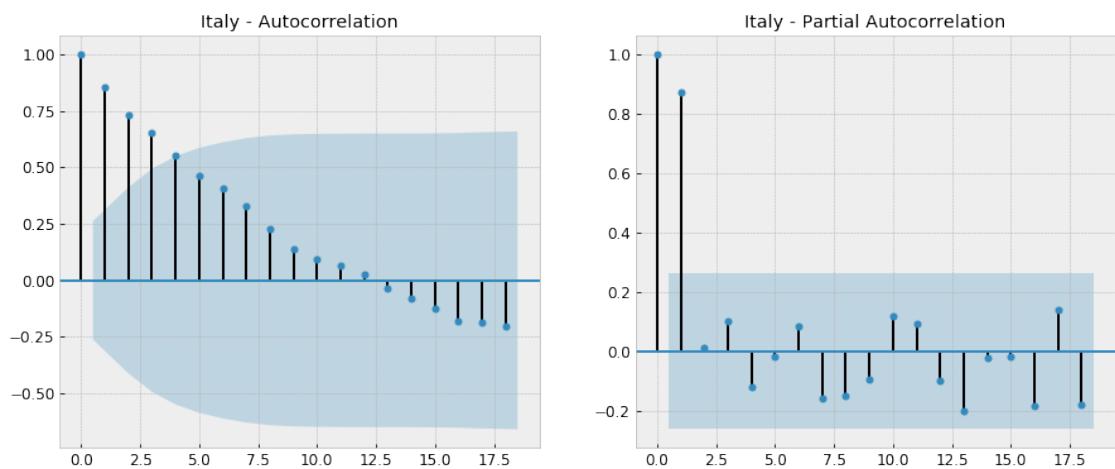
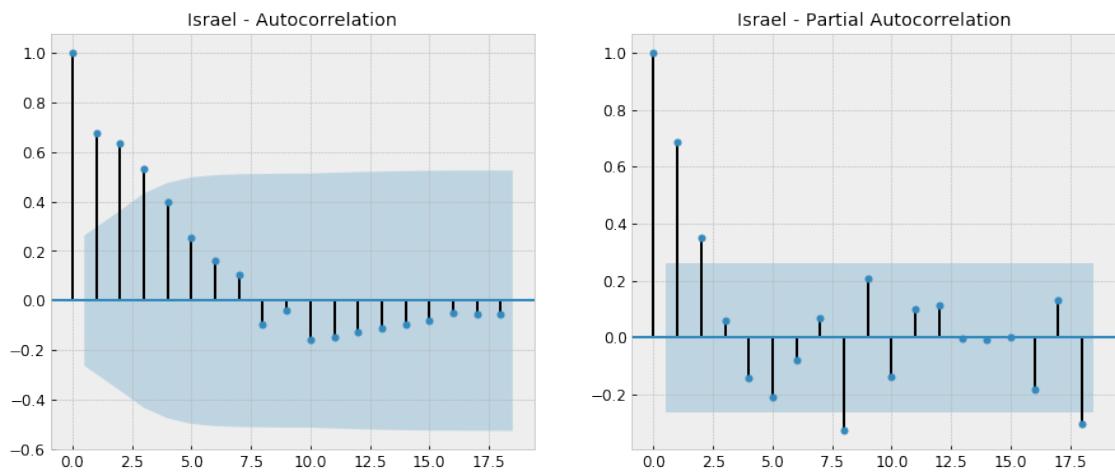


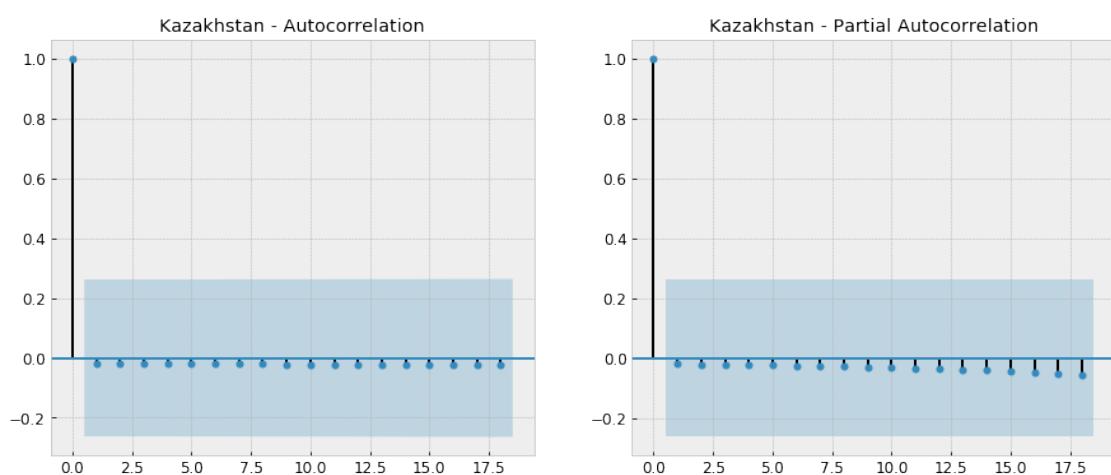
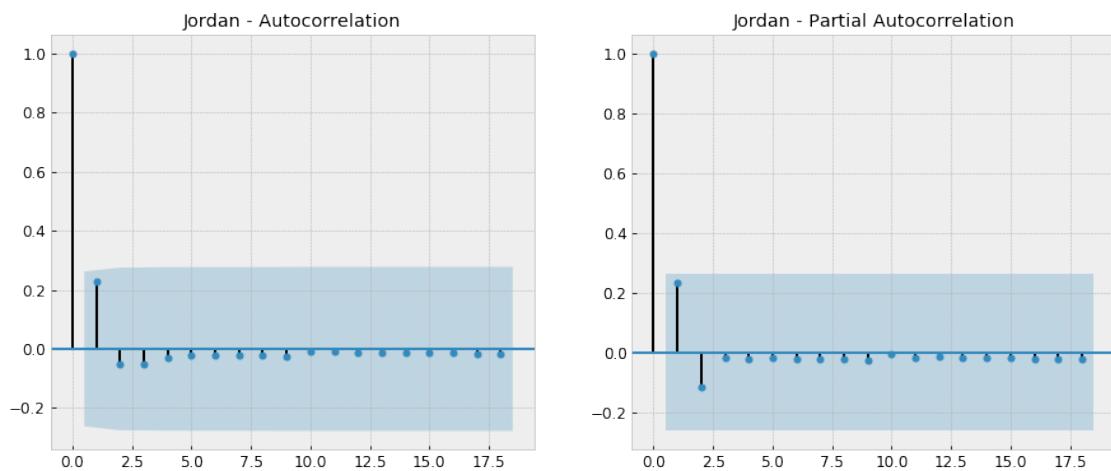


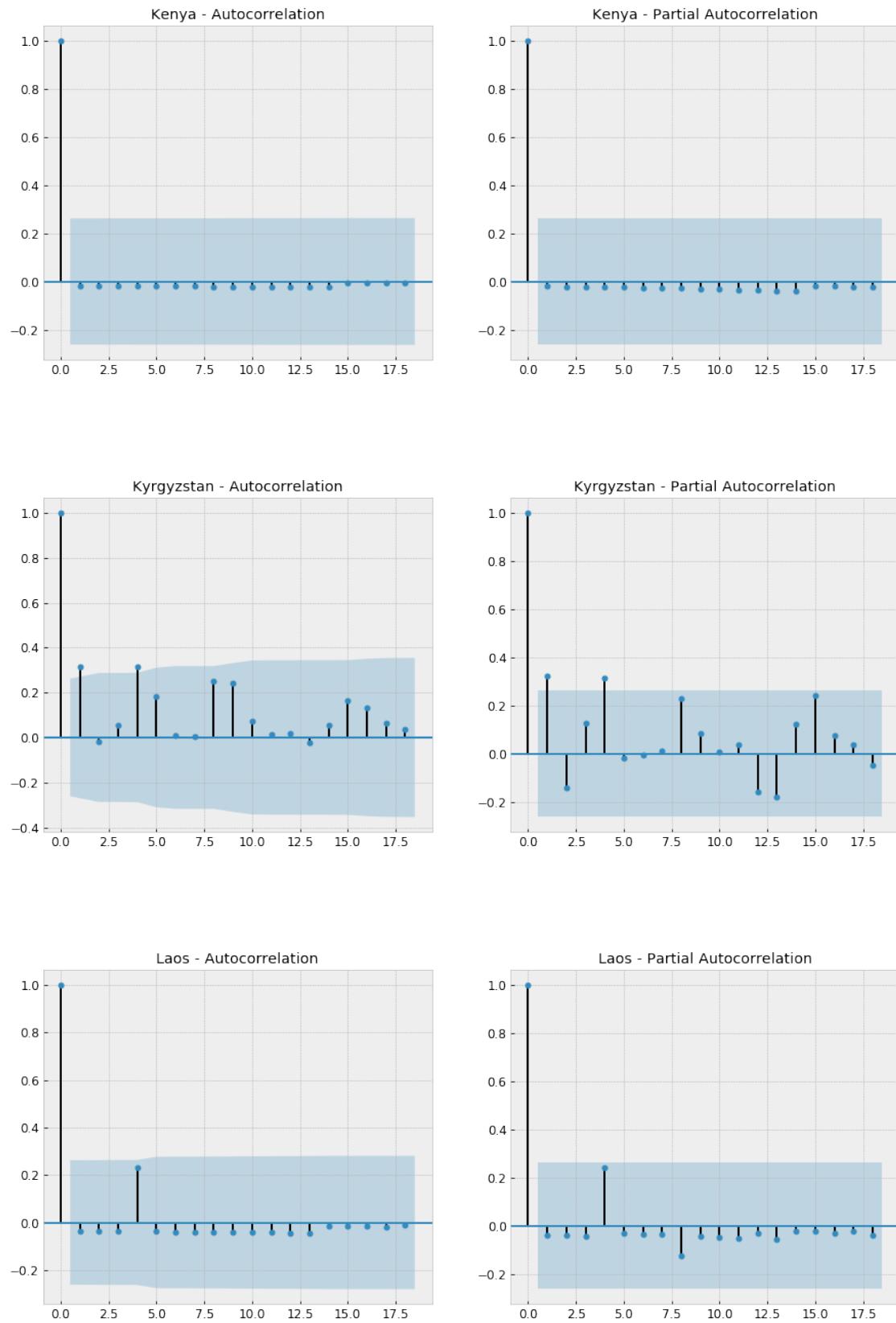


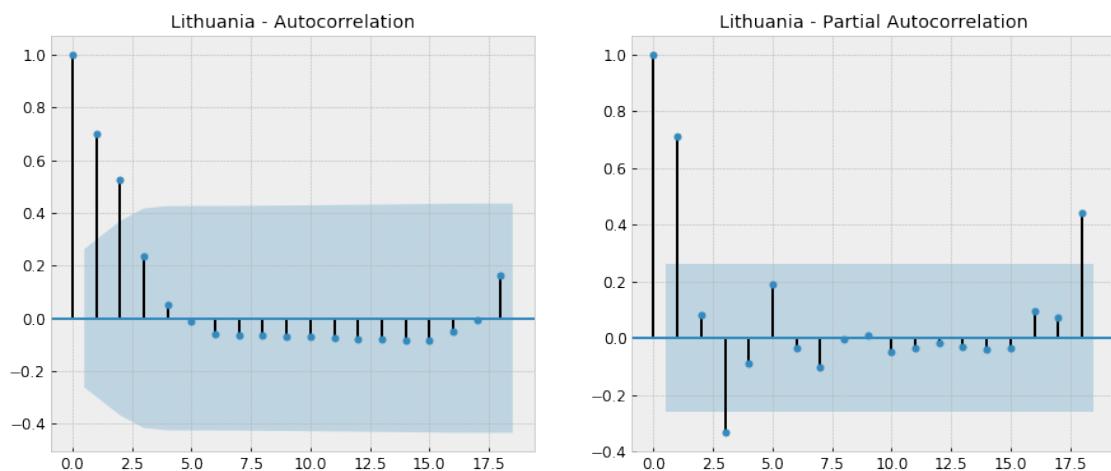
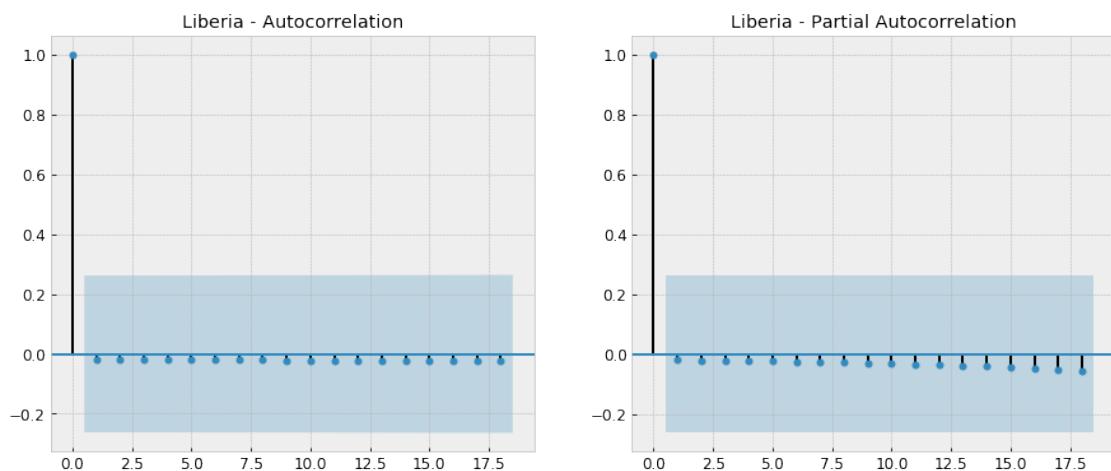
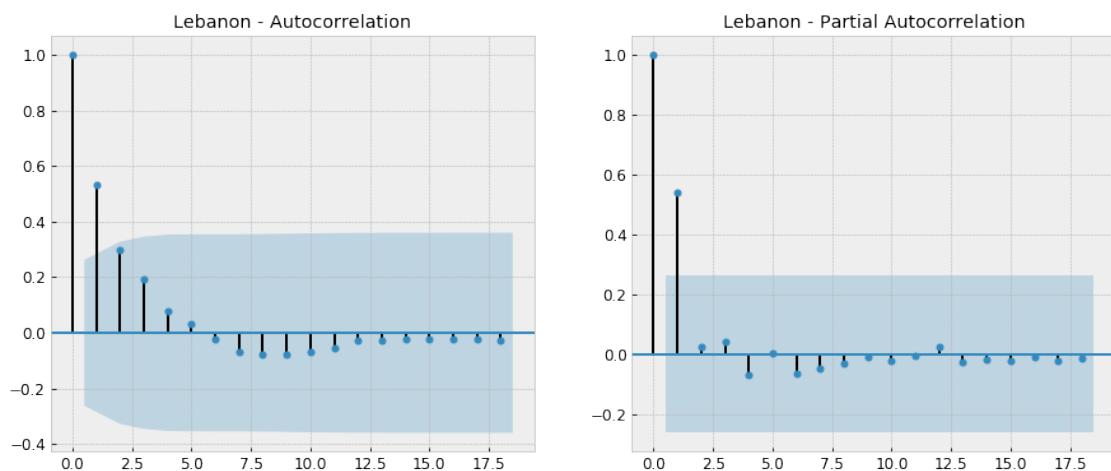


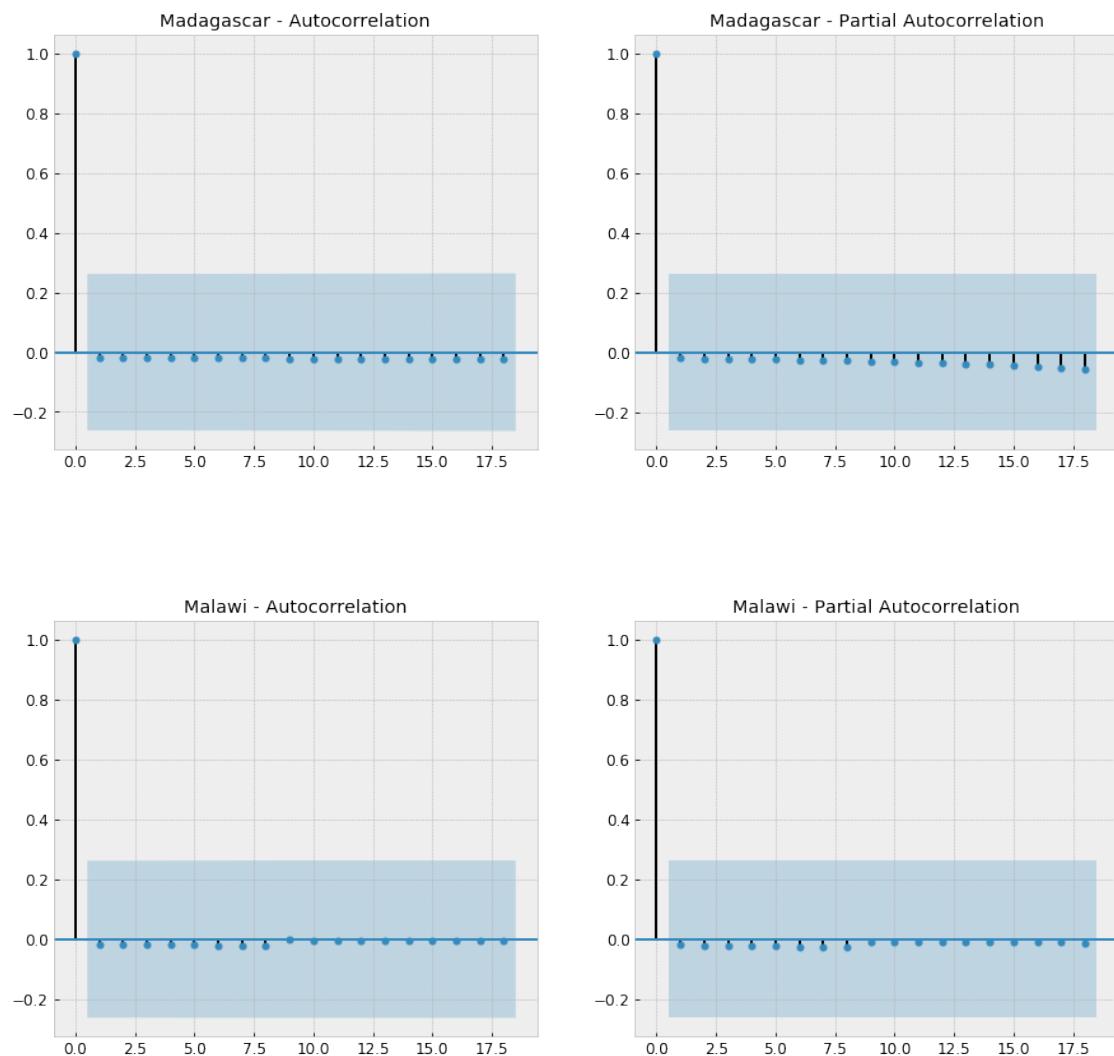




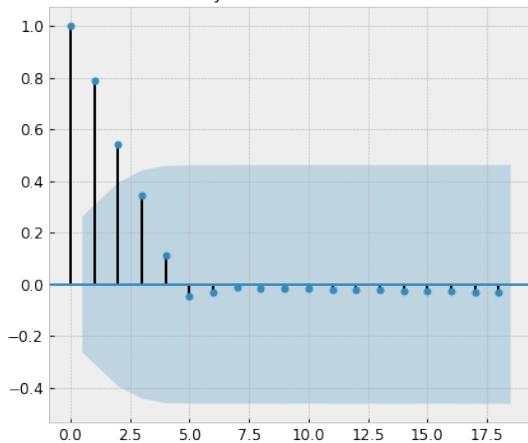




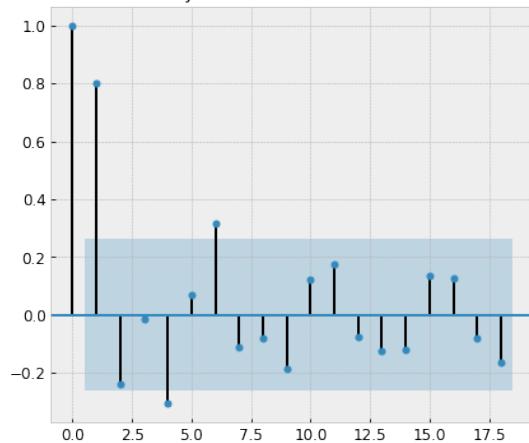




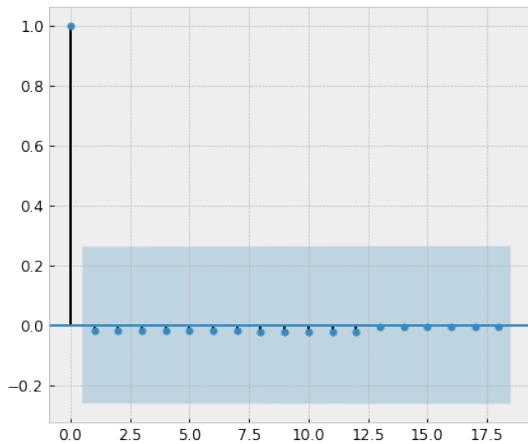
Malaysia - Autocorrelation



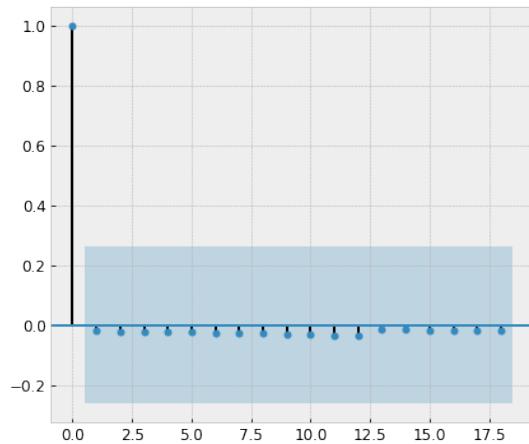
Malaysia - Partial Autocorrelation



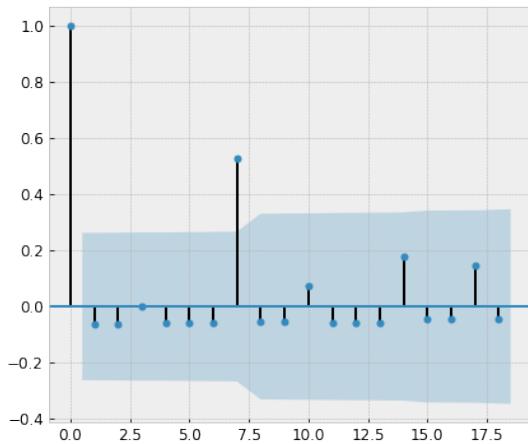
Maldives - Autocorrelation



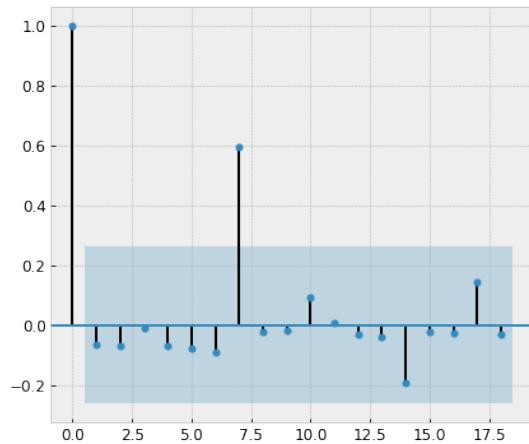
Maldives - Partial Autocorrelation

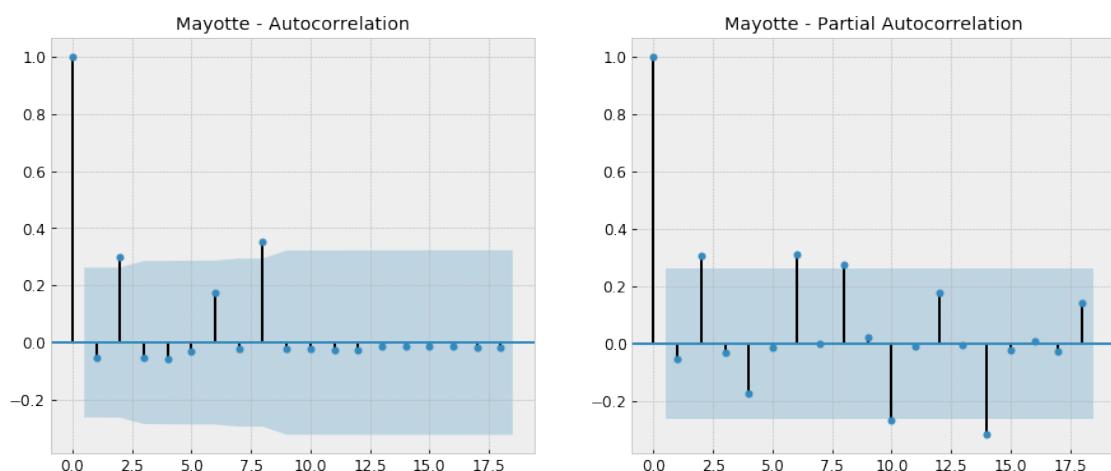
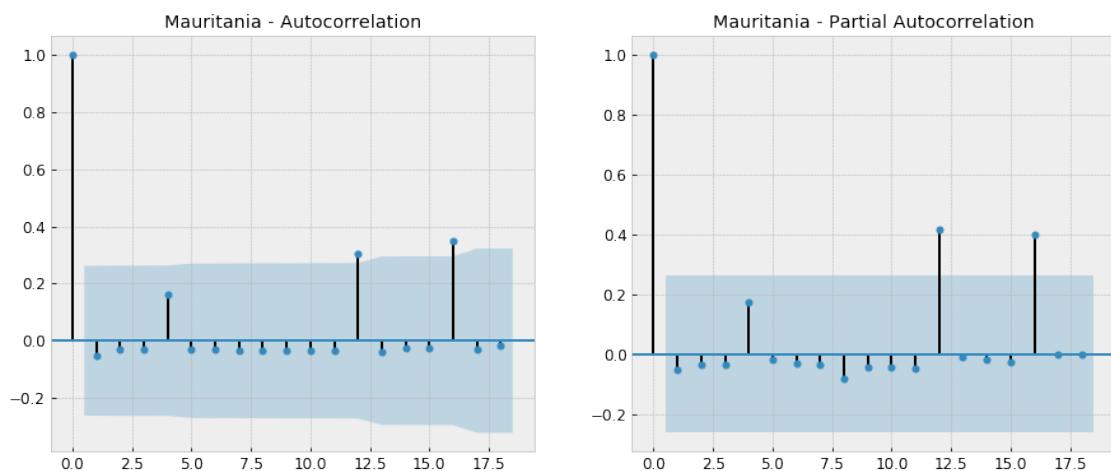
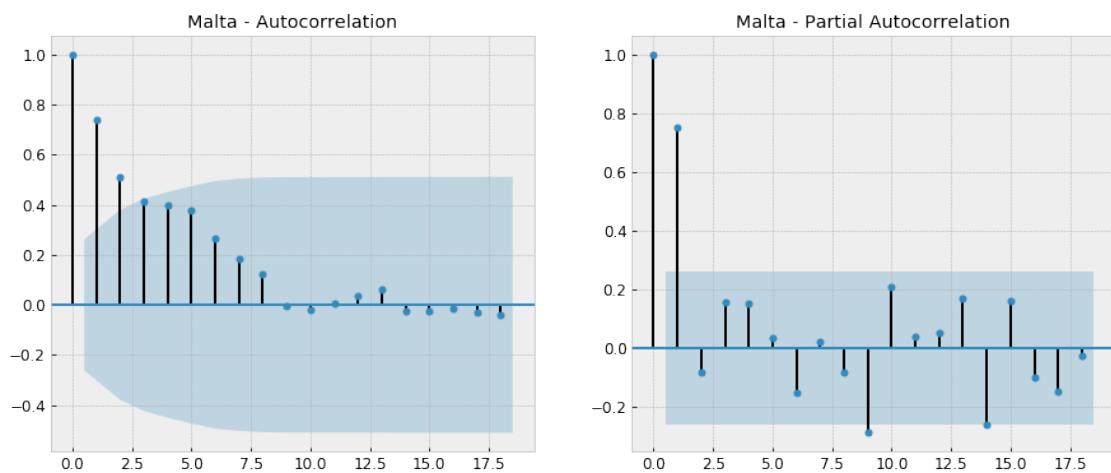


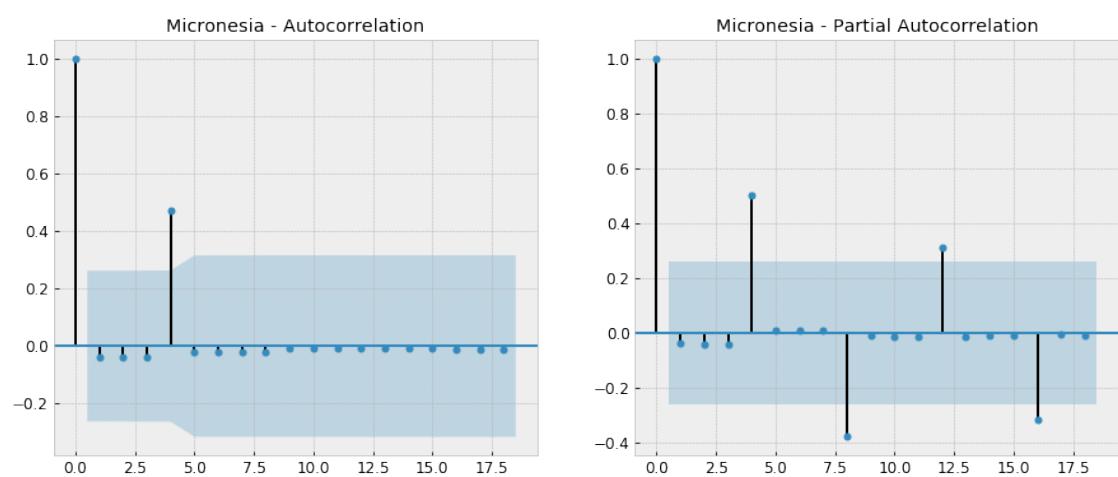
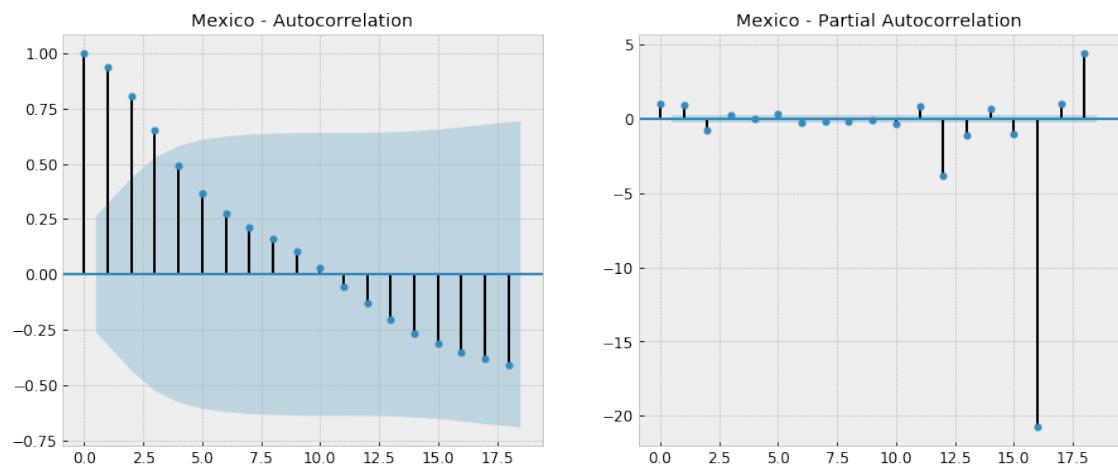
Mali - Autocorrelation

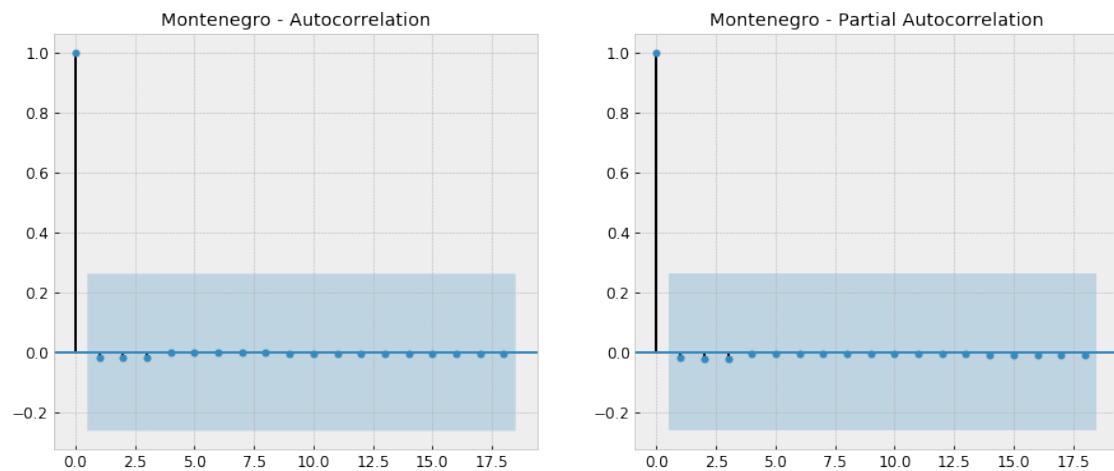
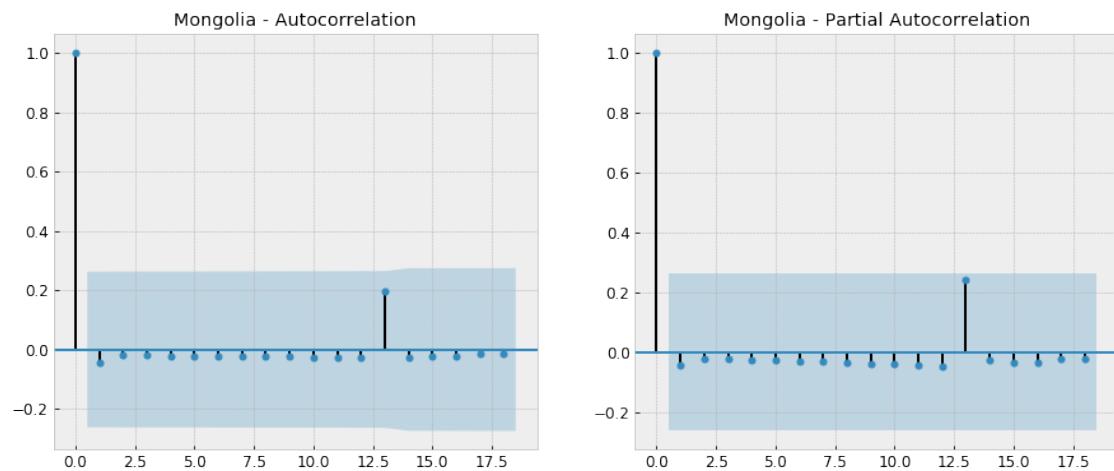
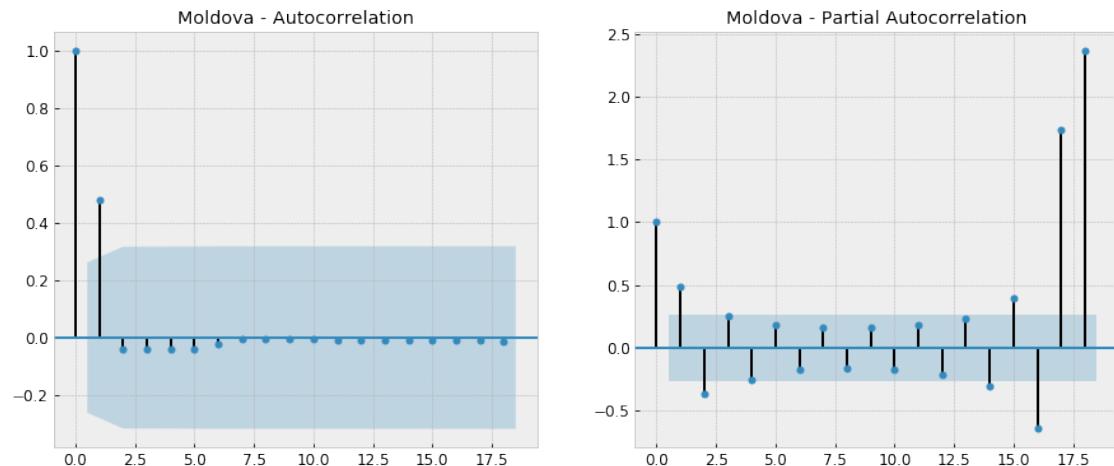


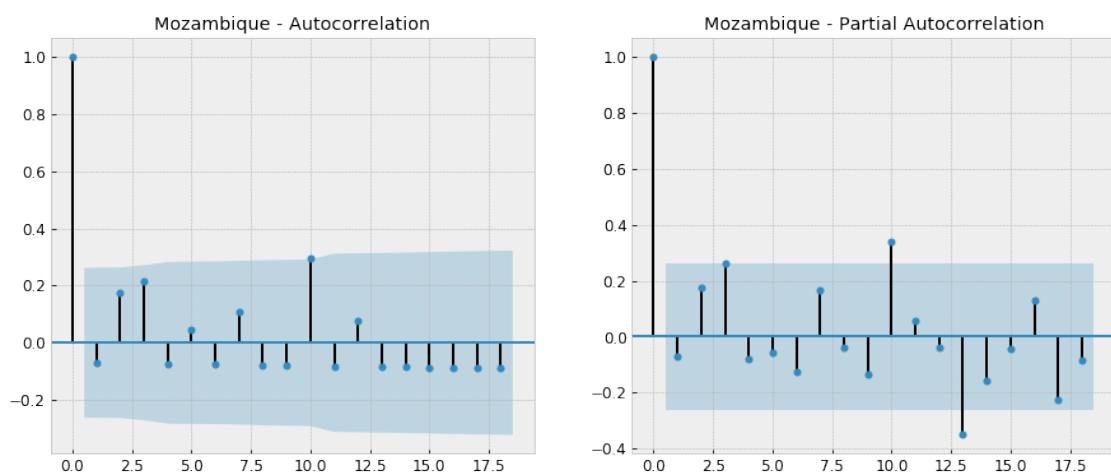
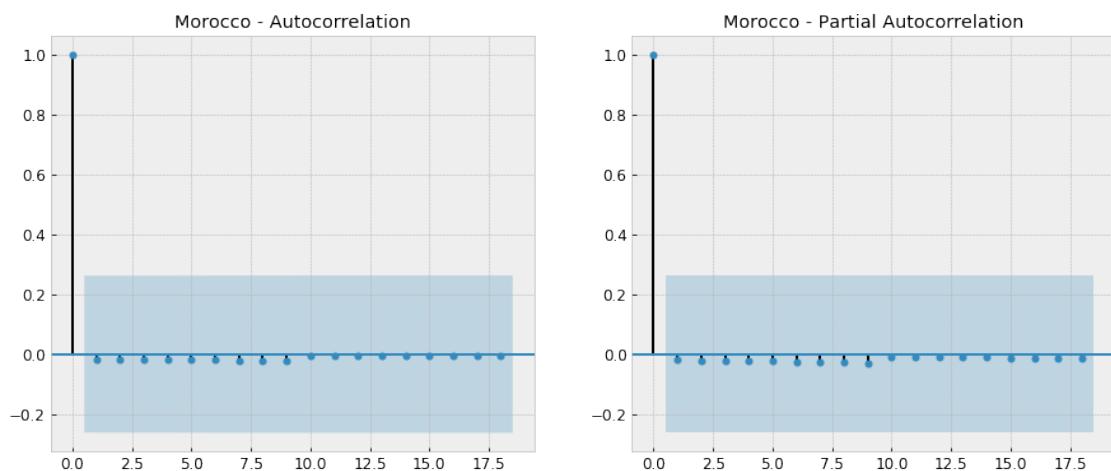
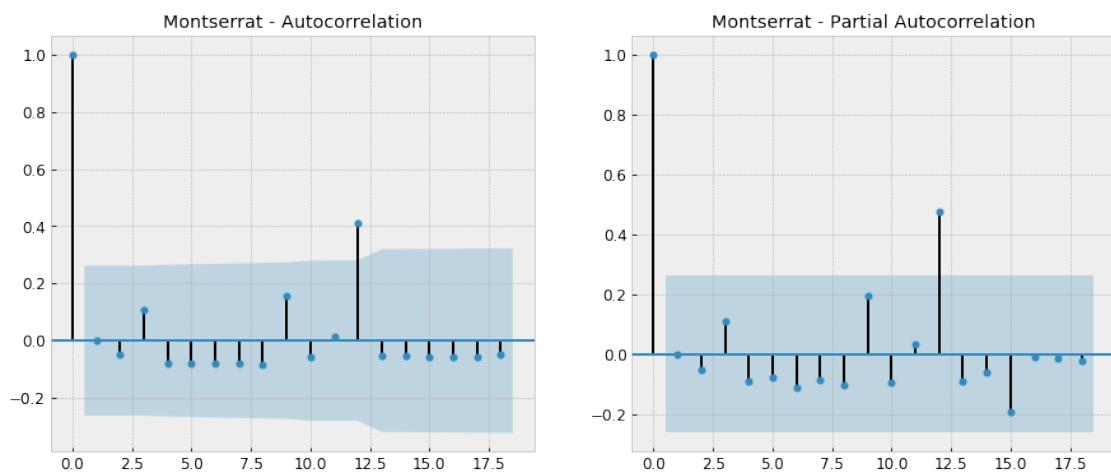
Mali - Partial Autocorrelation

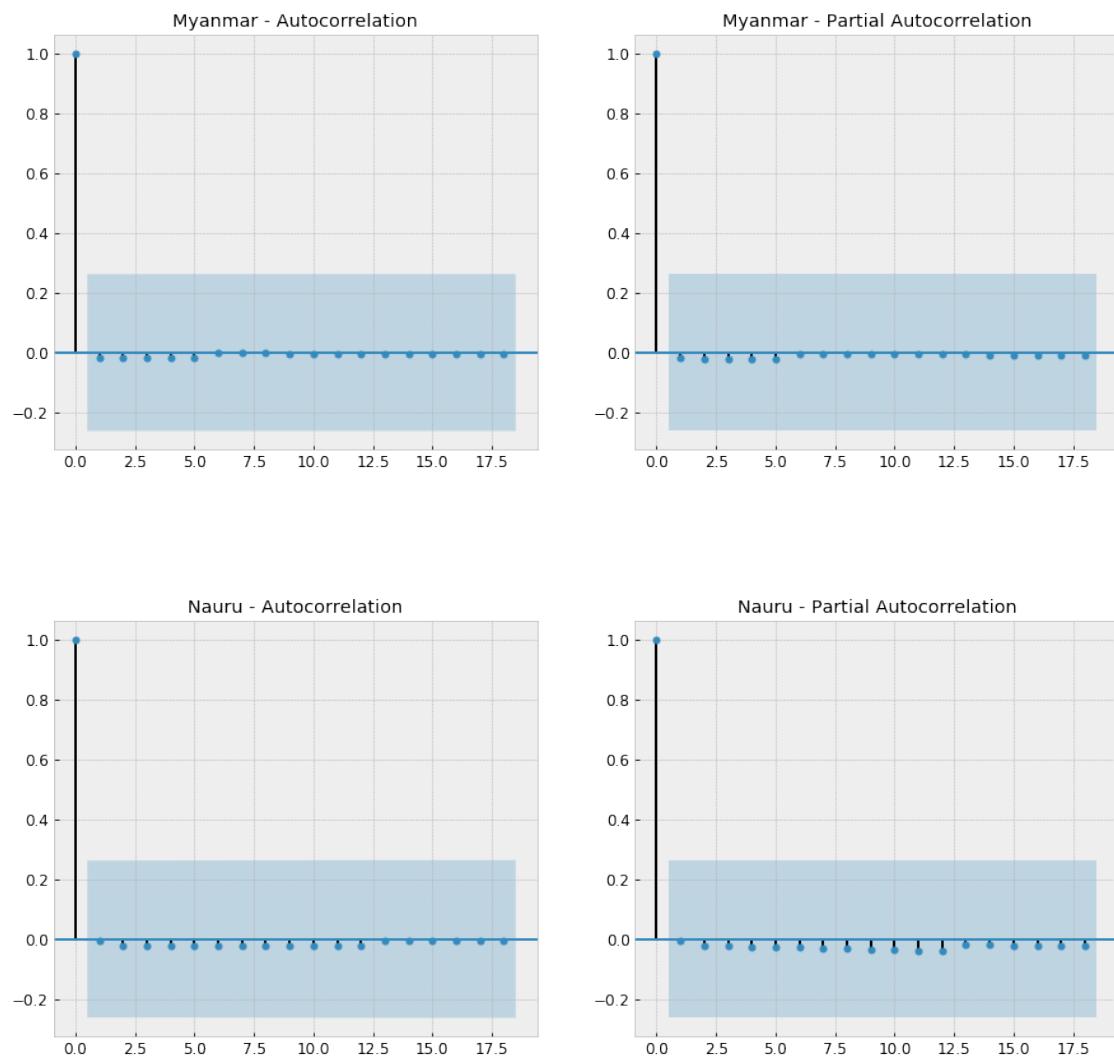




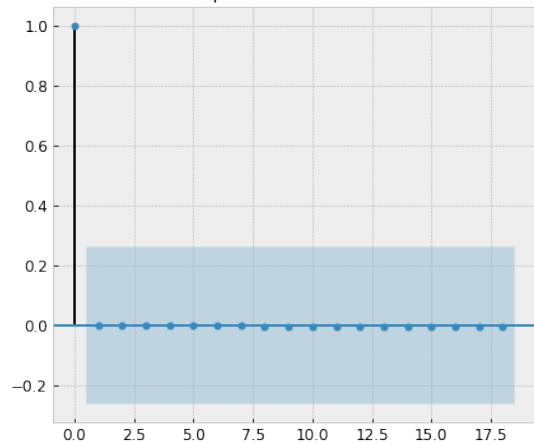




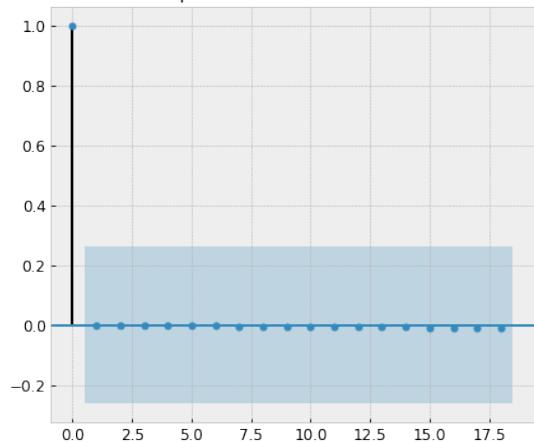




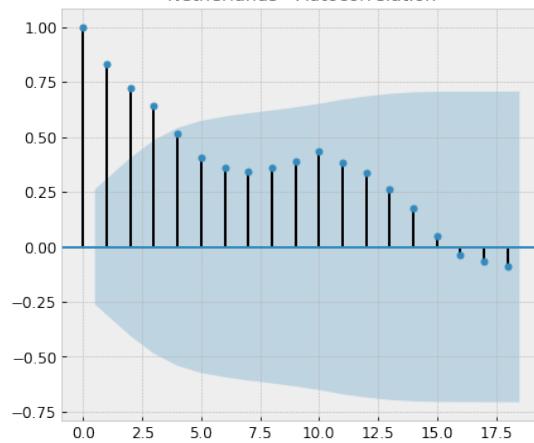
Nepal - Autocorrelation



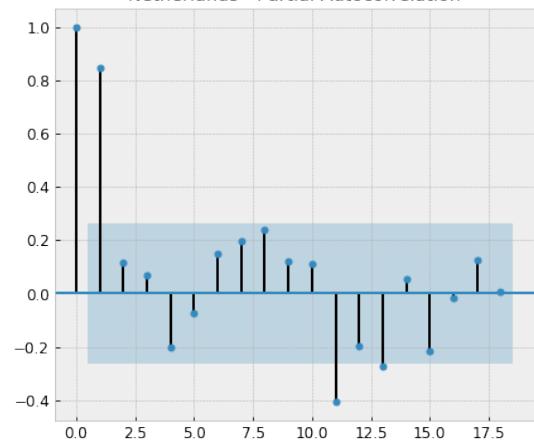
Nepal - Partial Autocorrelation



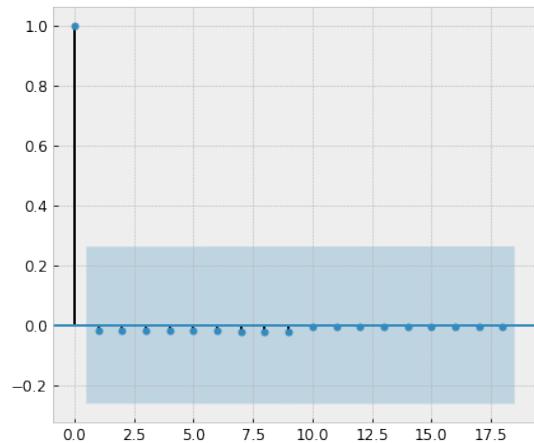
Netherlands - Autocorrelation



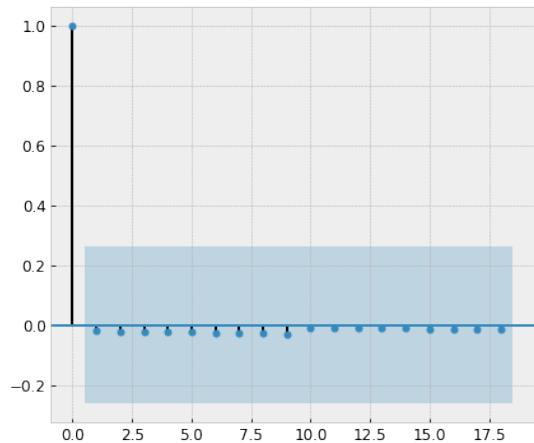
Netherlands - Partial Autocorrelation

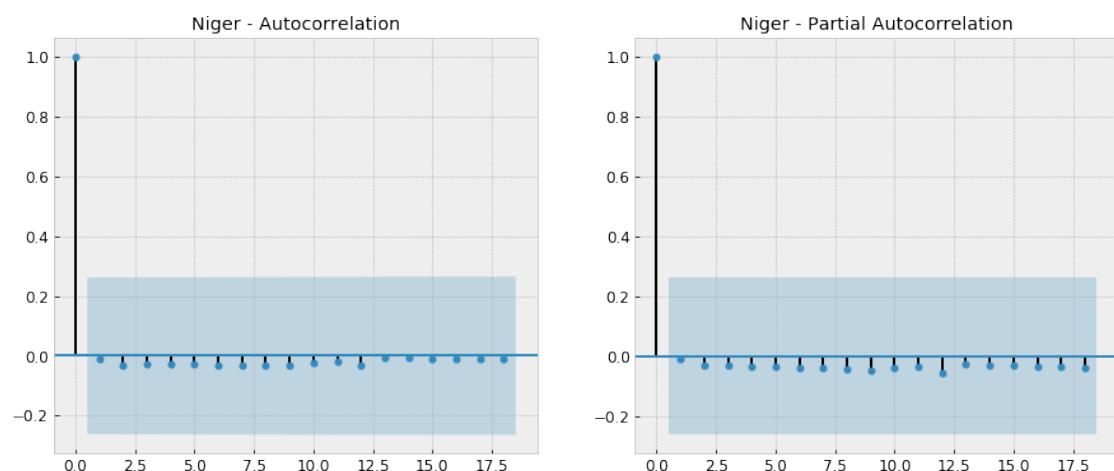
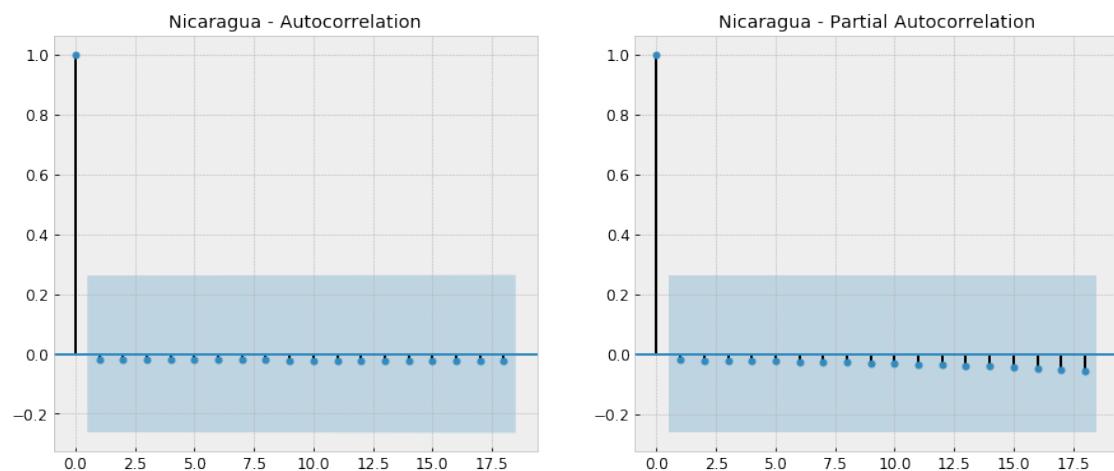
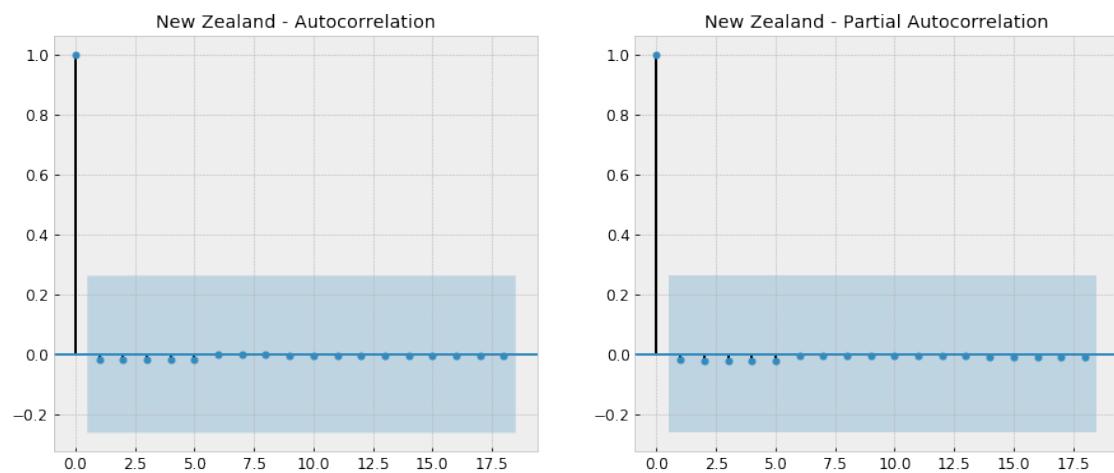


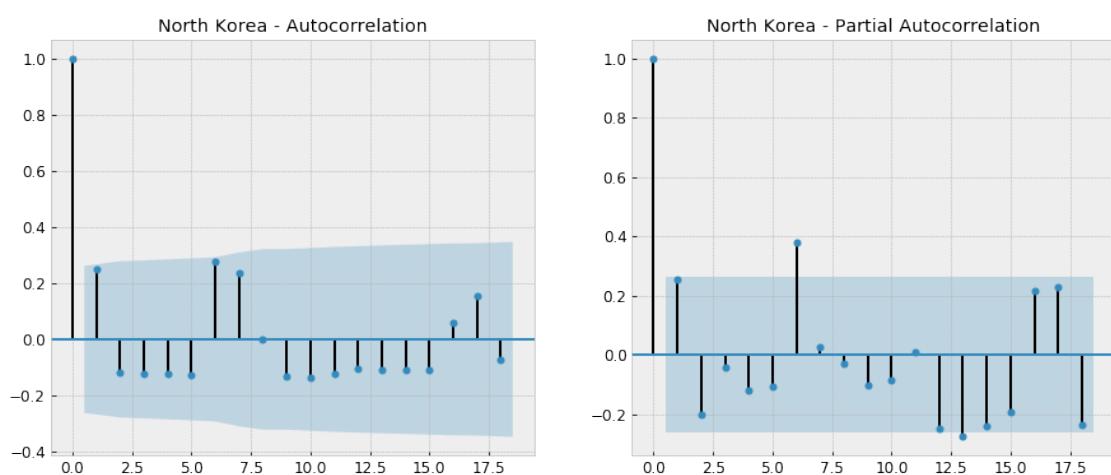
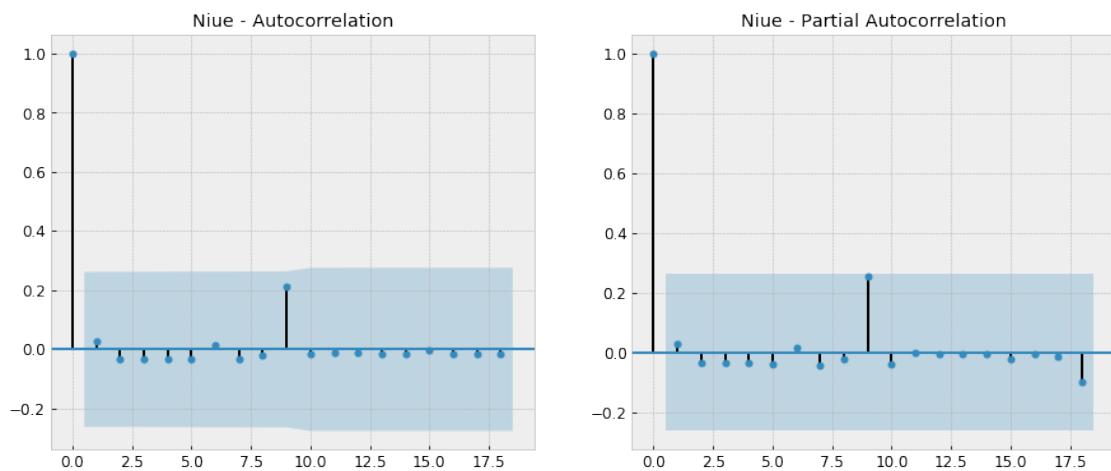
Netherlands Antilles - Autocorrelation

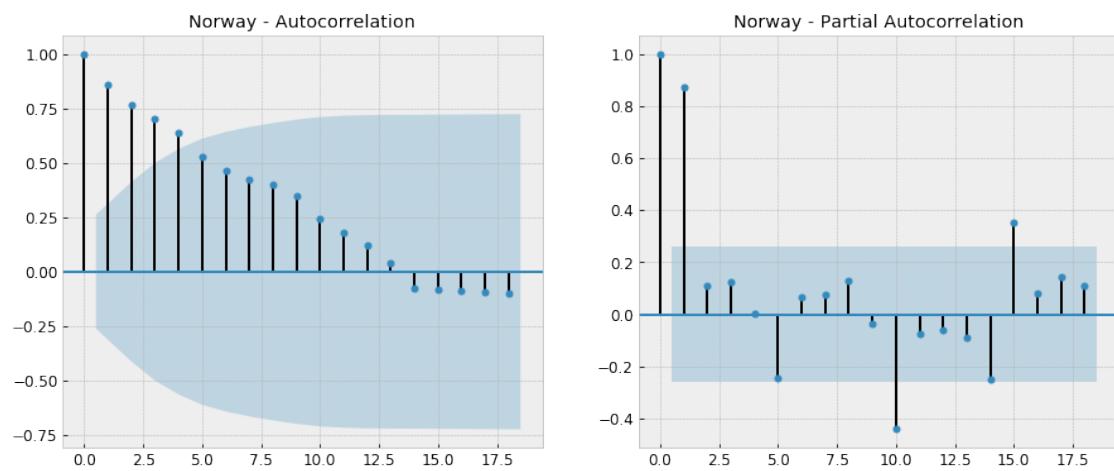
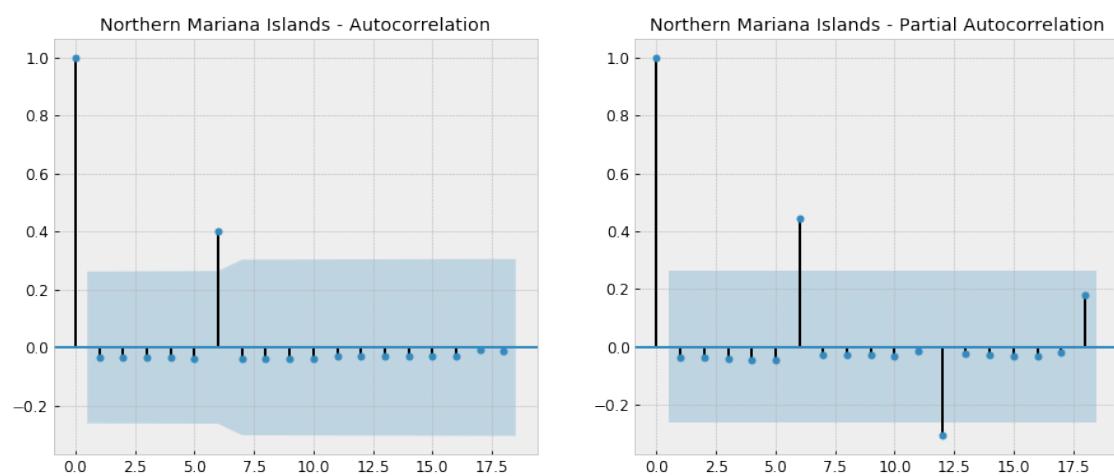
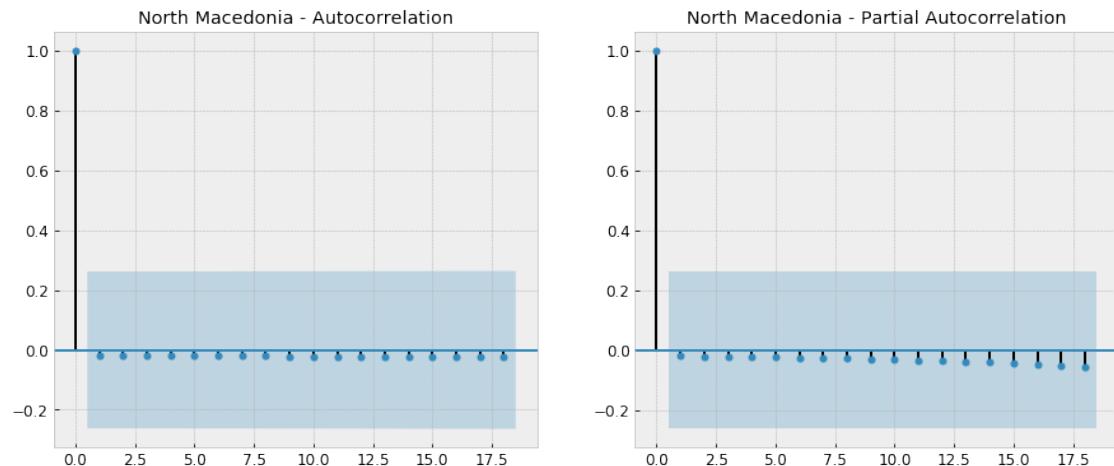


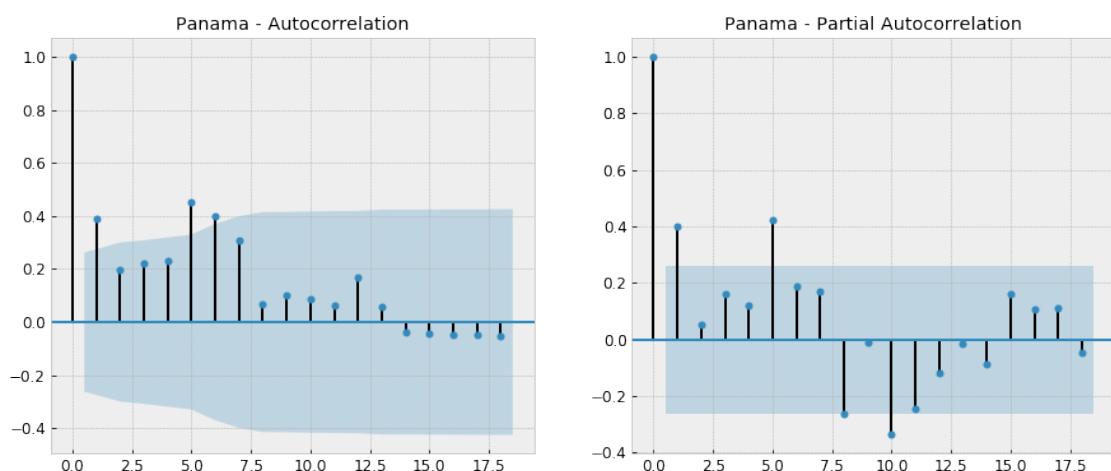
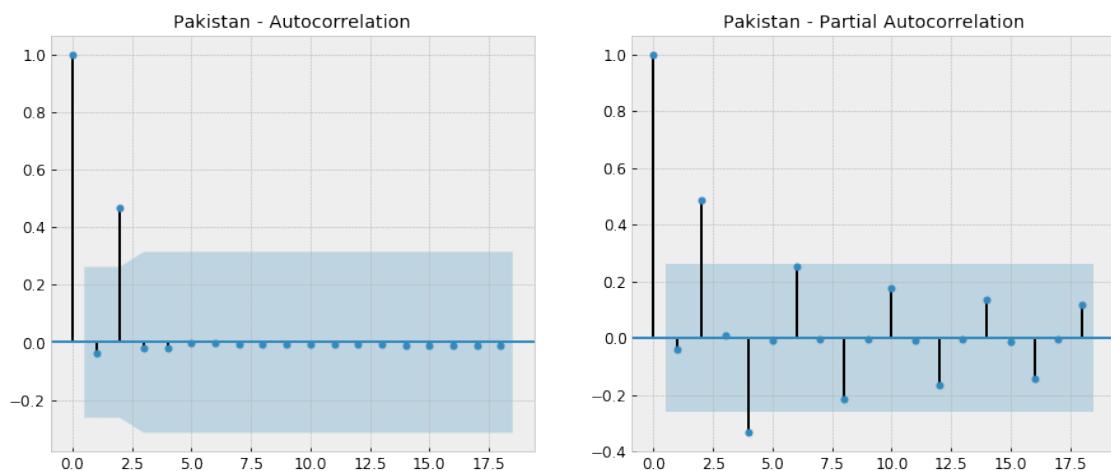
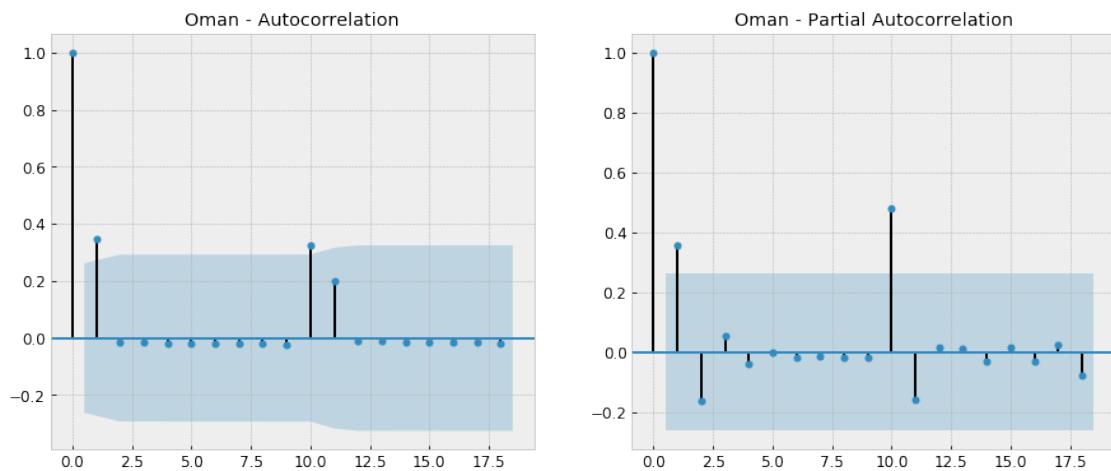
Netherlands Antilles - Partial Autocorrelation

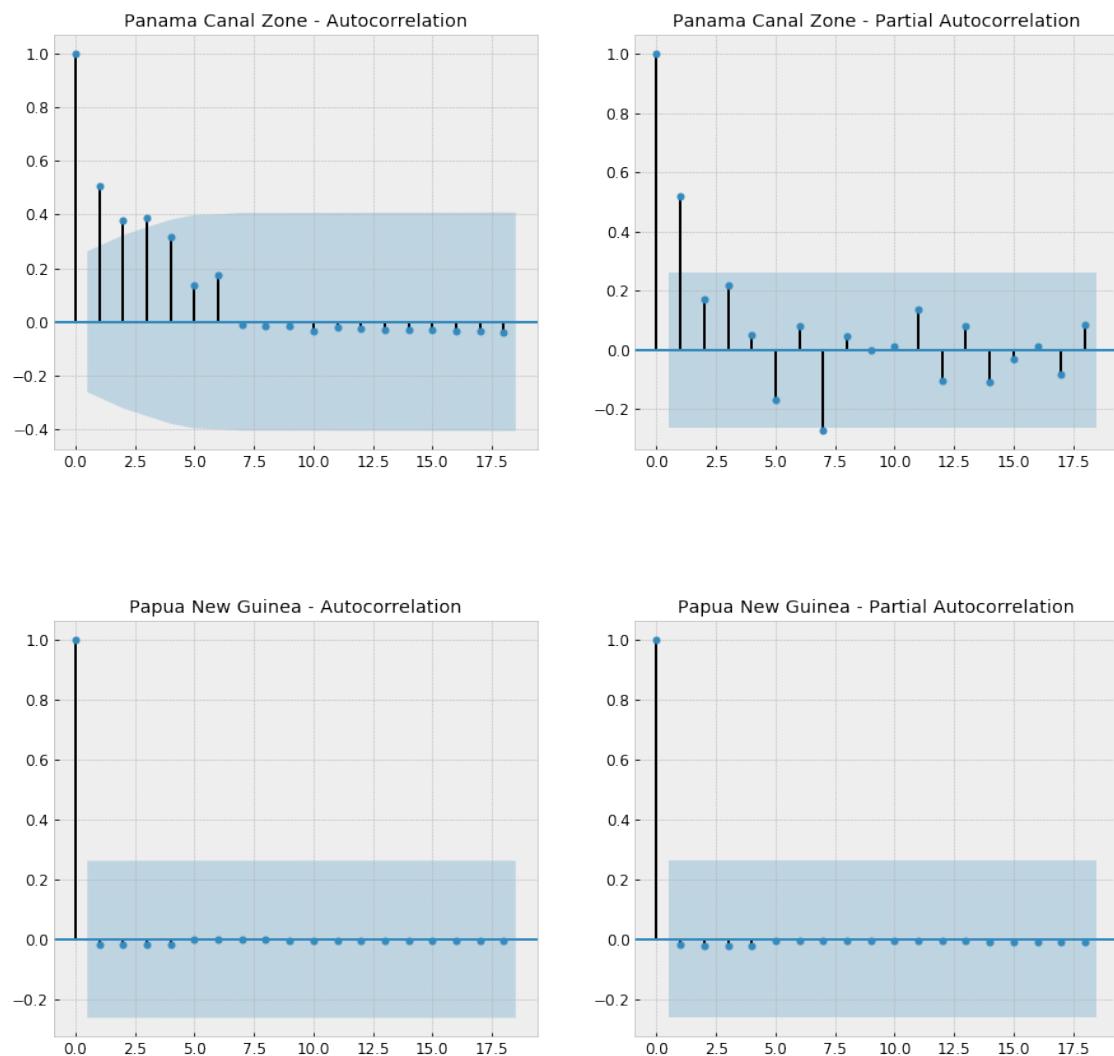


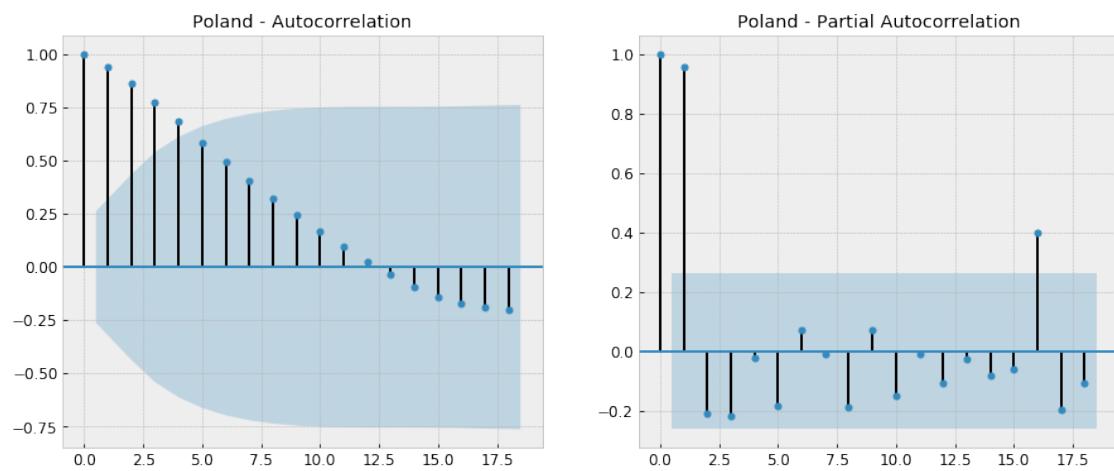
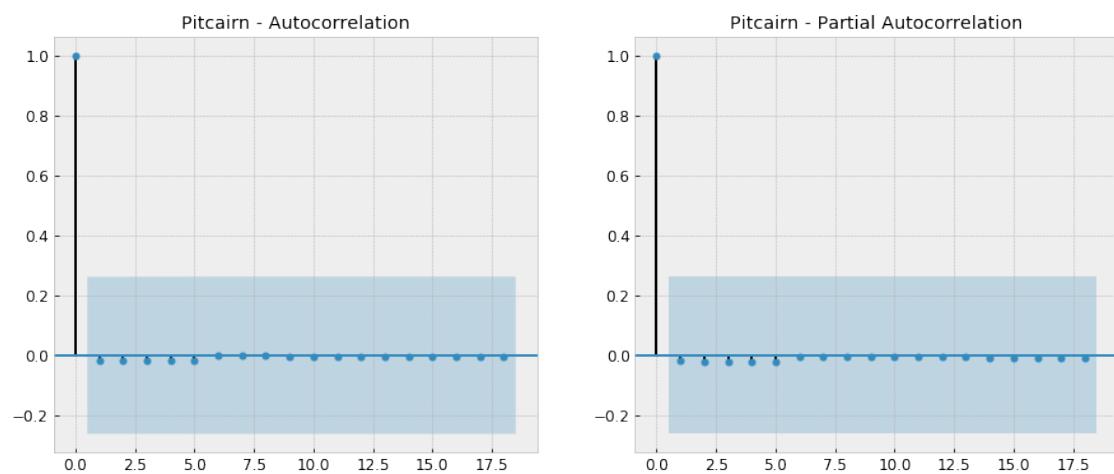
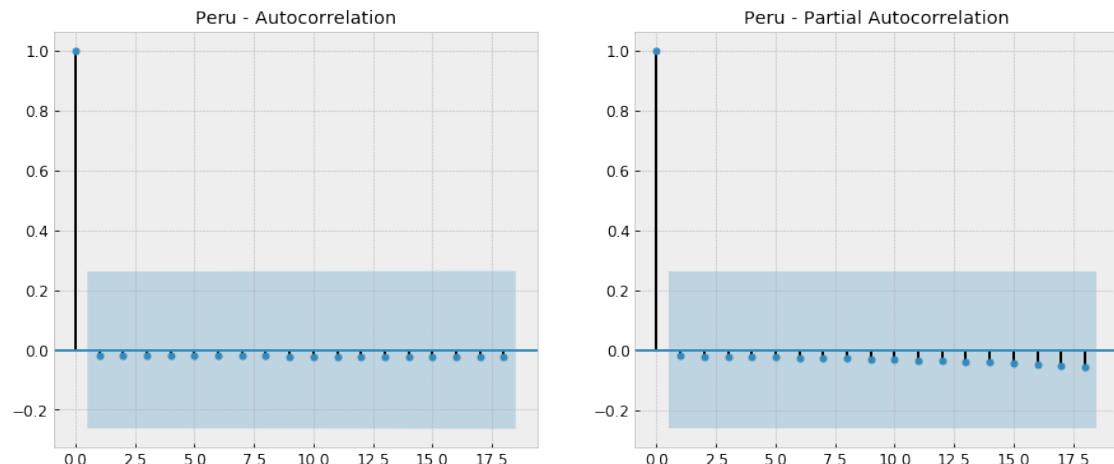


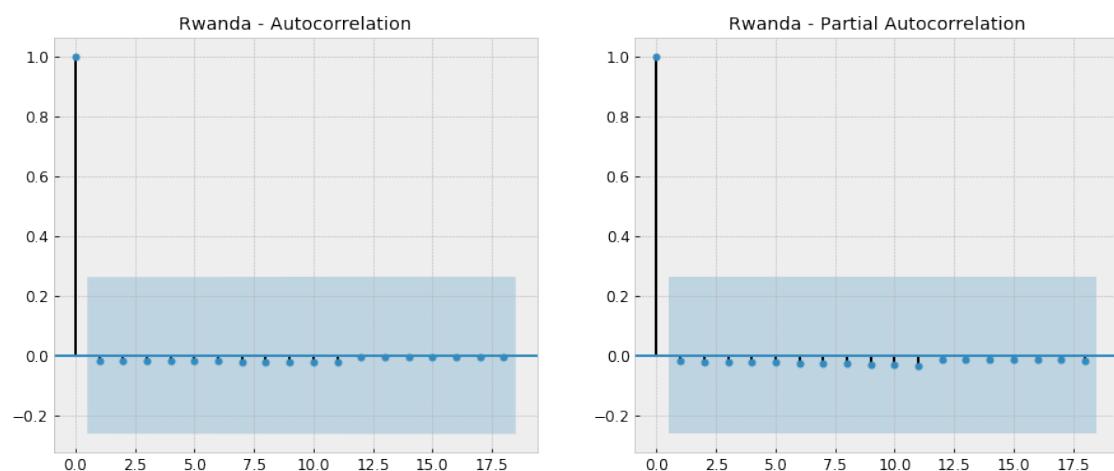
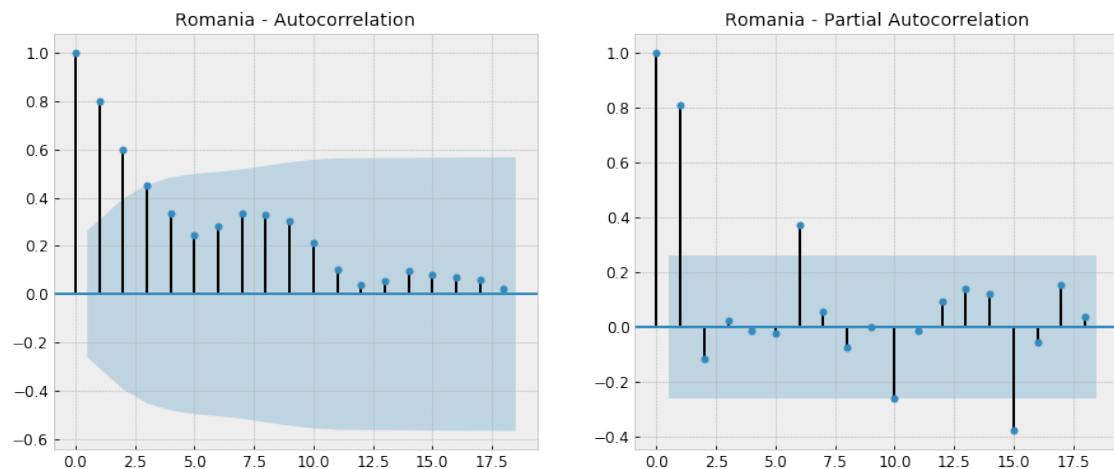
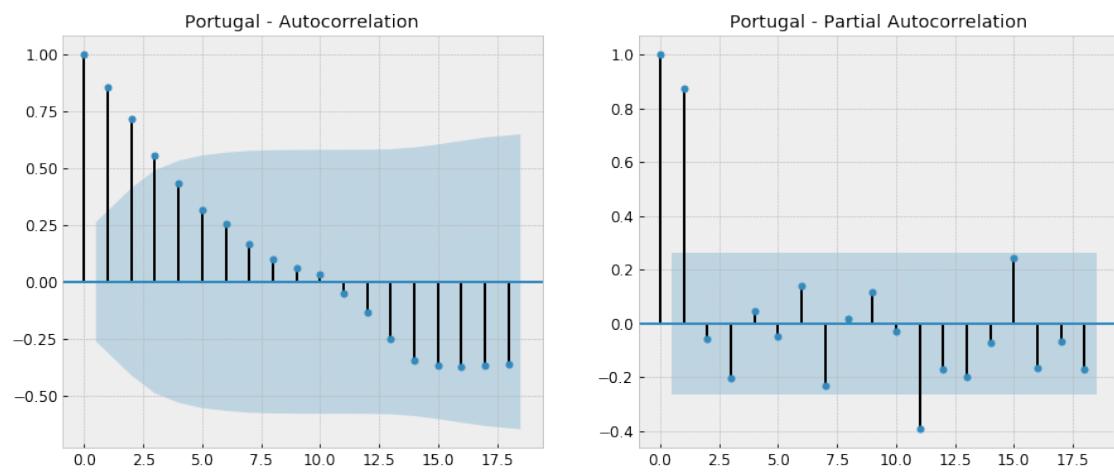




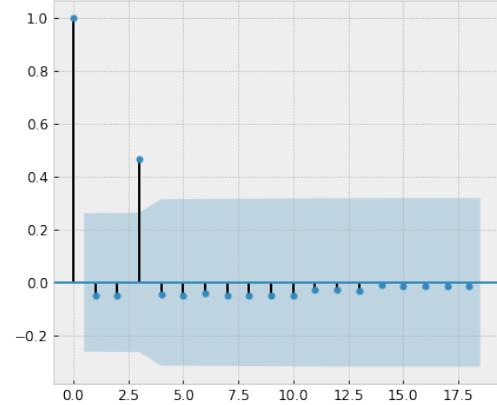




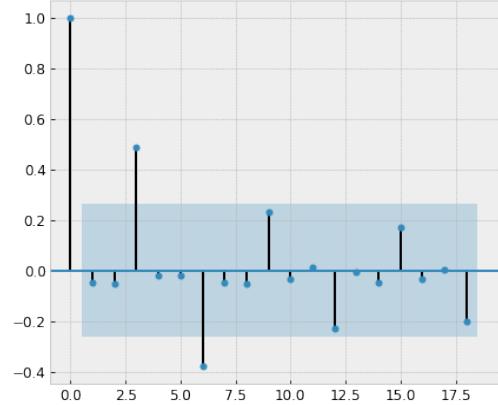




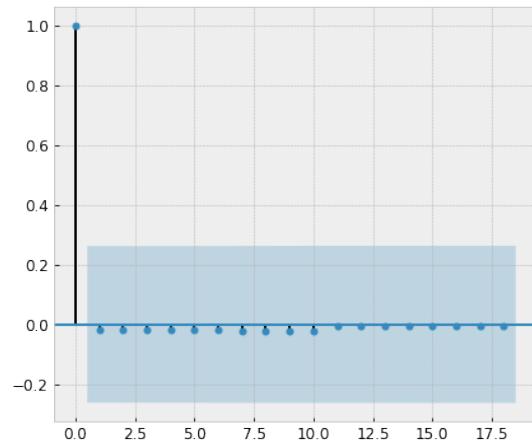
Saint Helena, Ascension and Tristan da Cunha - Autocorrelation



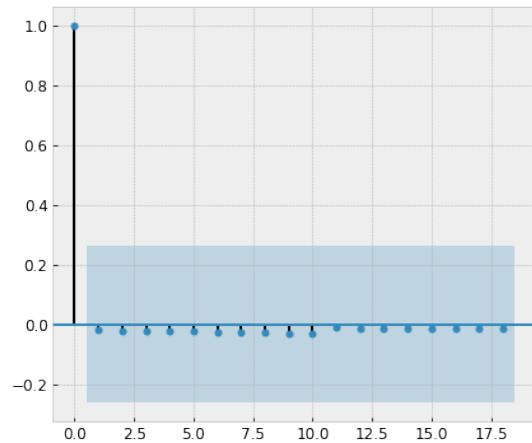
Saint Helena, Ascension and Tristan da Cunha - Partial Autocorrelation



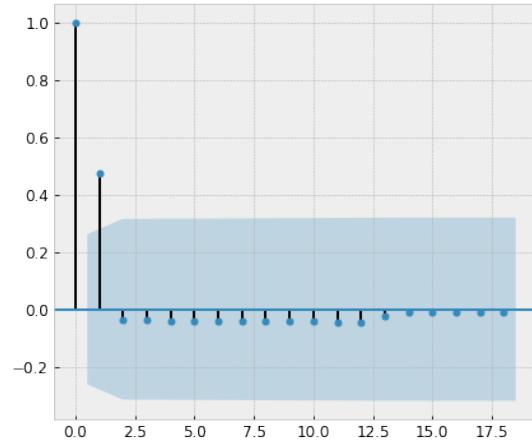
Saint Kitts and Nevis - Autocorrelation



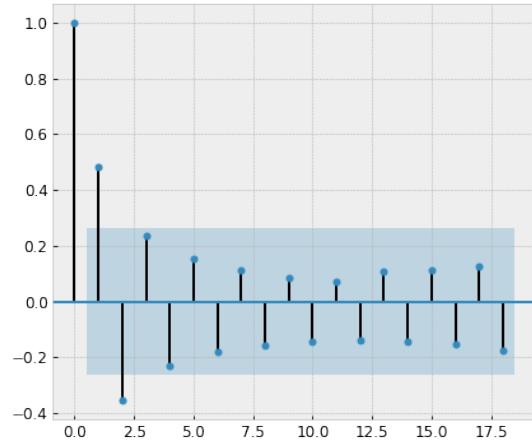
Saint Kitts and Nevis - Partial Autocorrelation

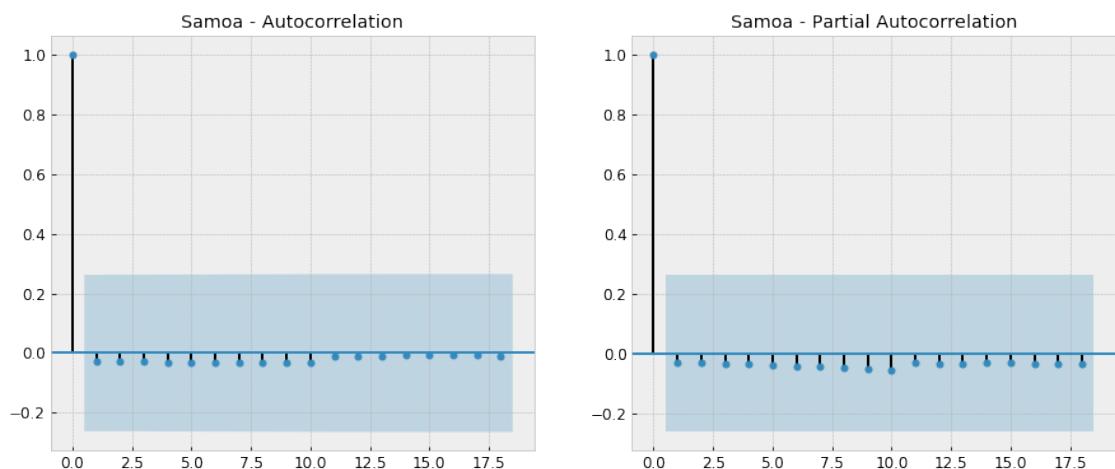
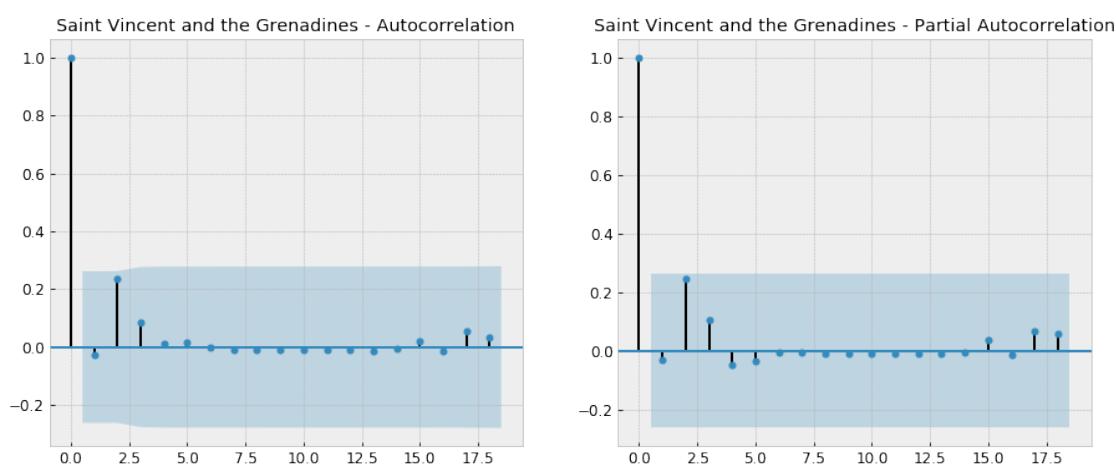
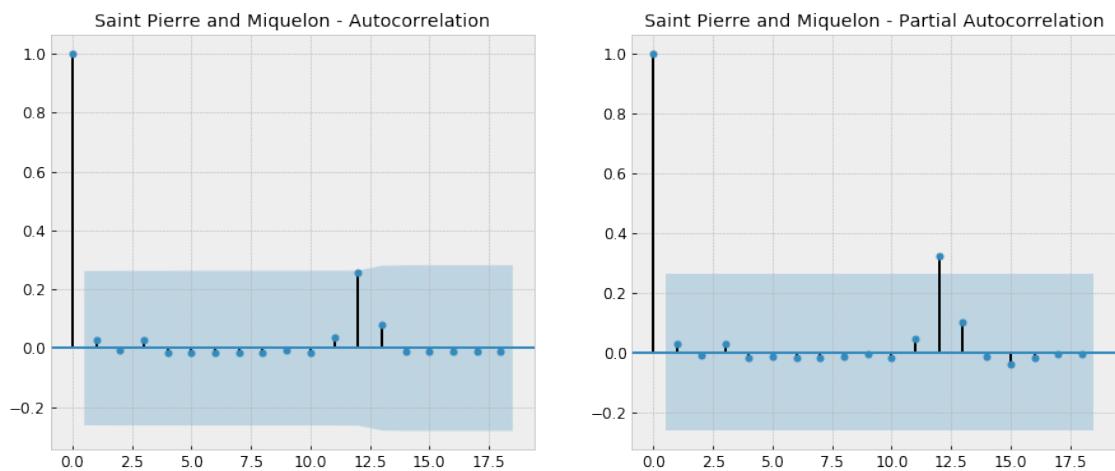


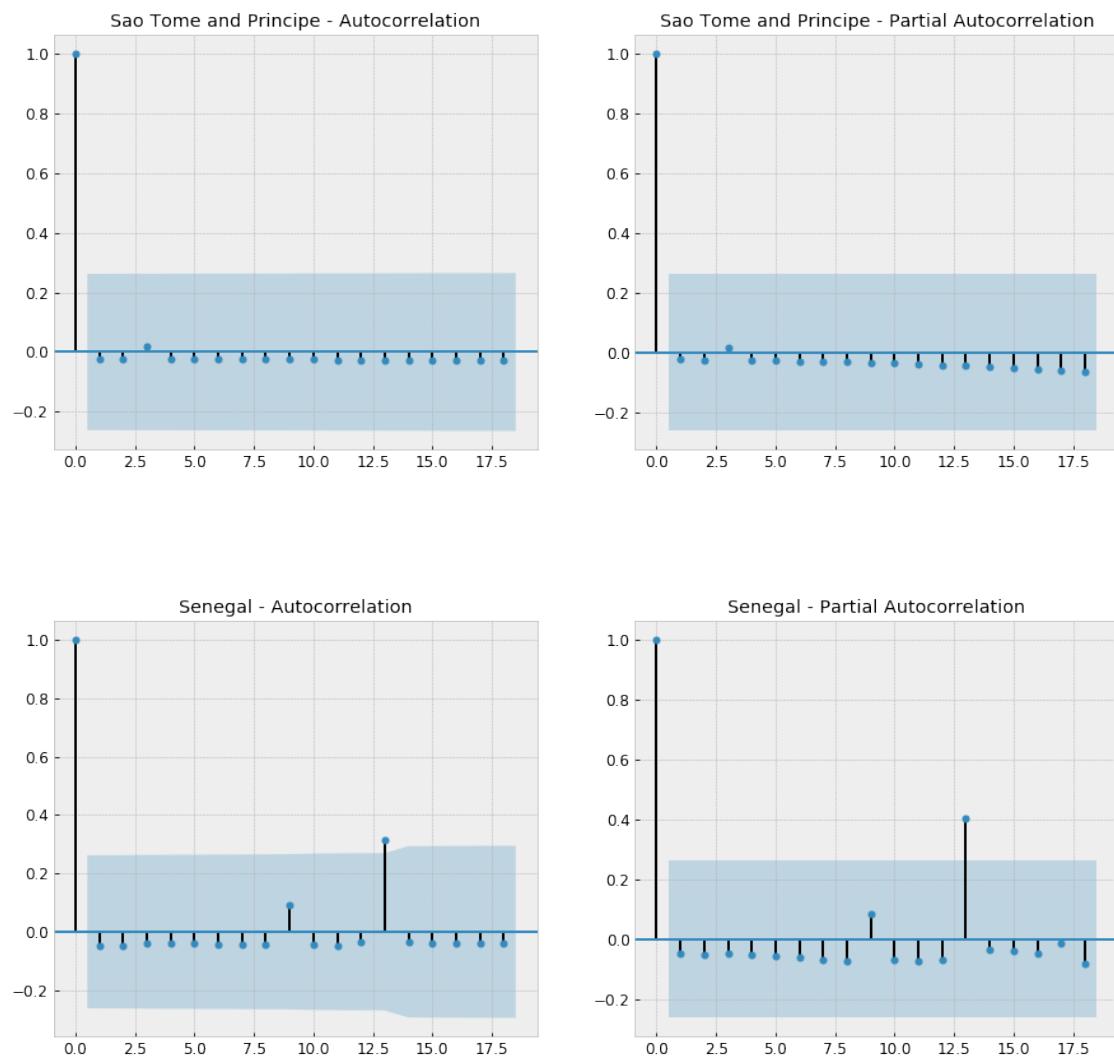
Saint Lucia - Autocorrelation

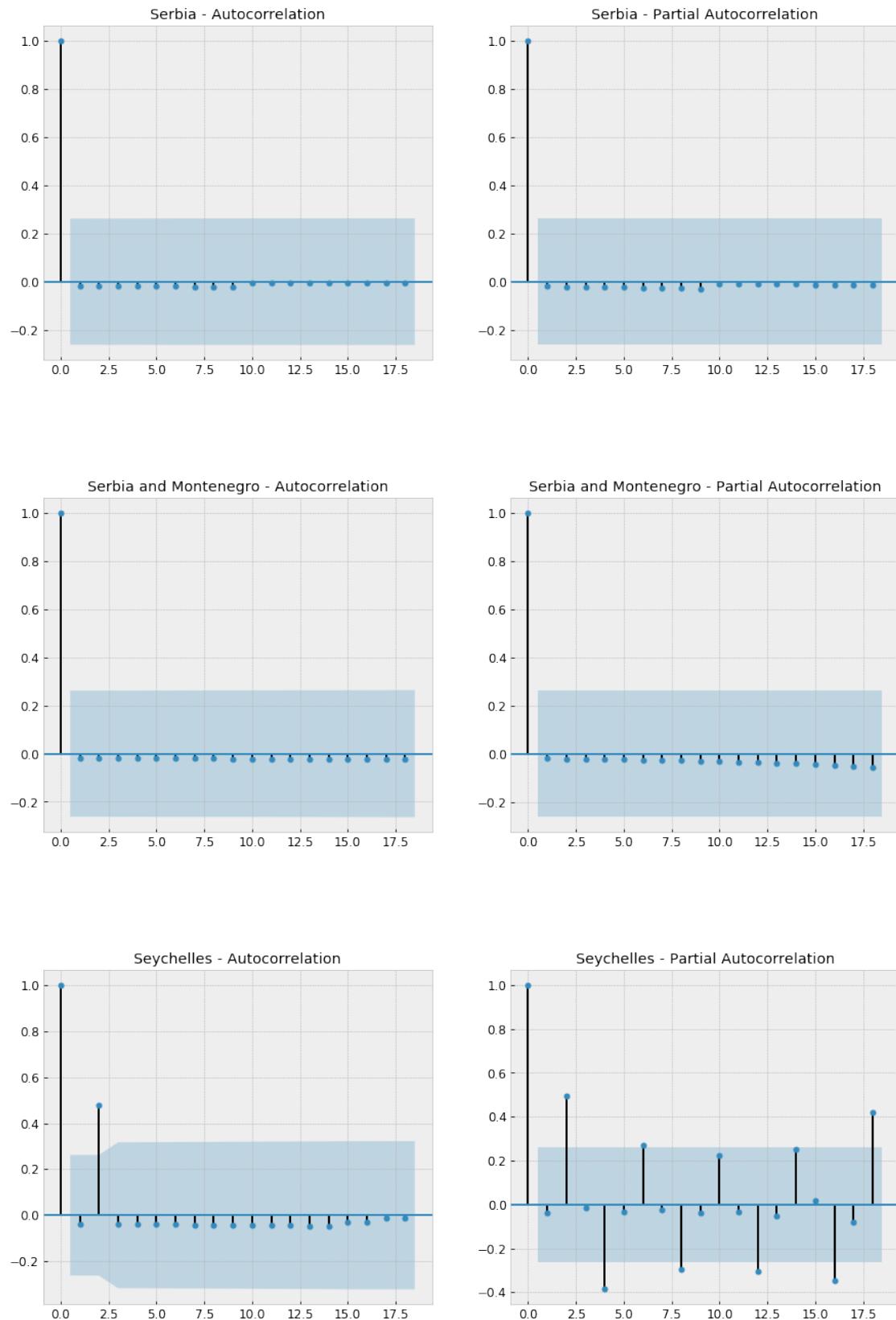


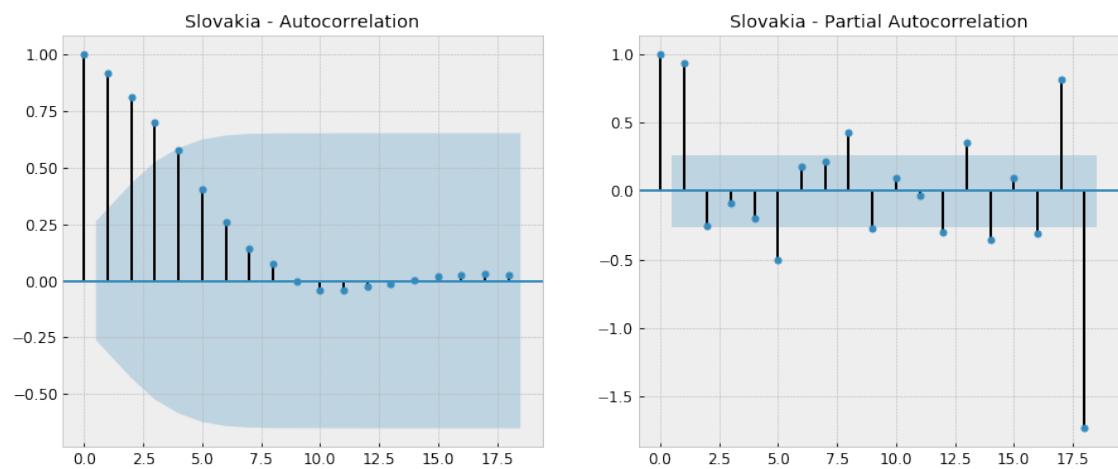
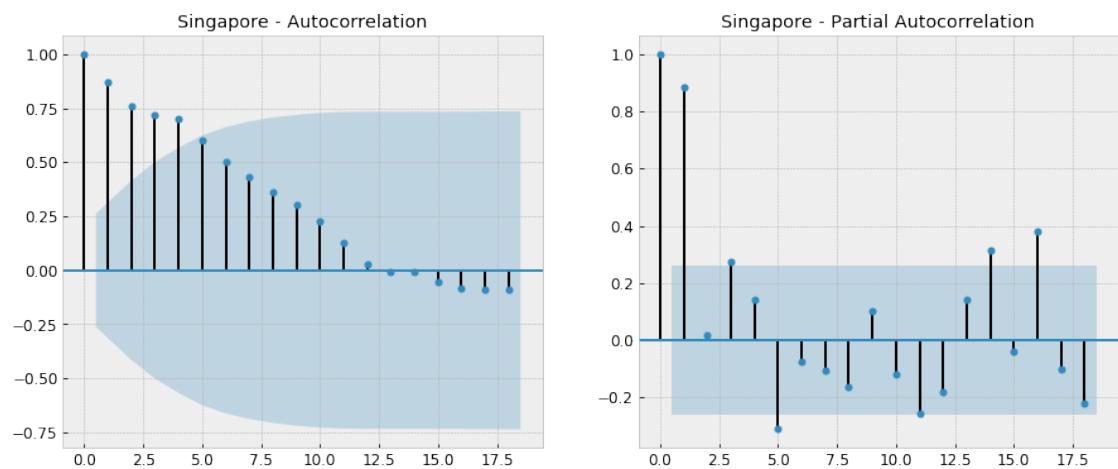
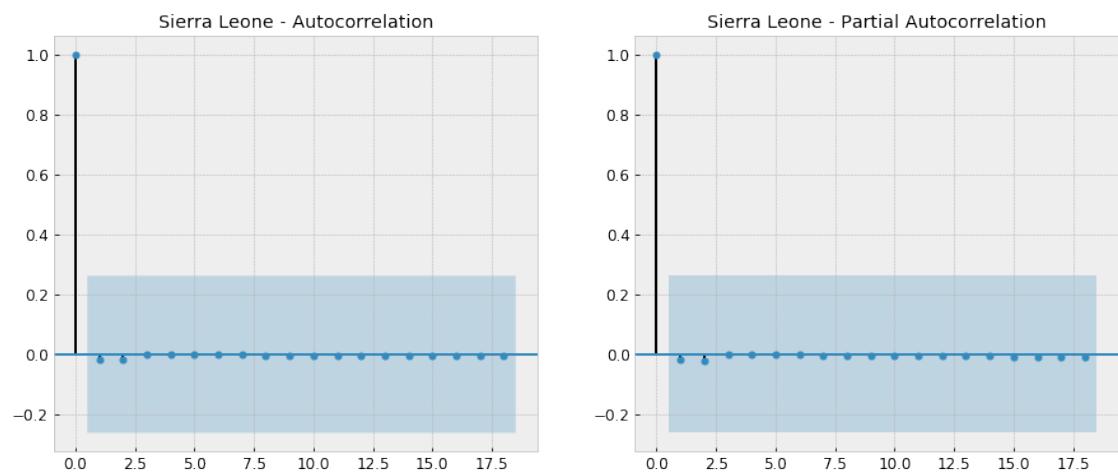
Saint Lucia - Partial Autocorrelation

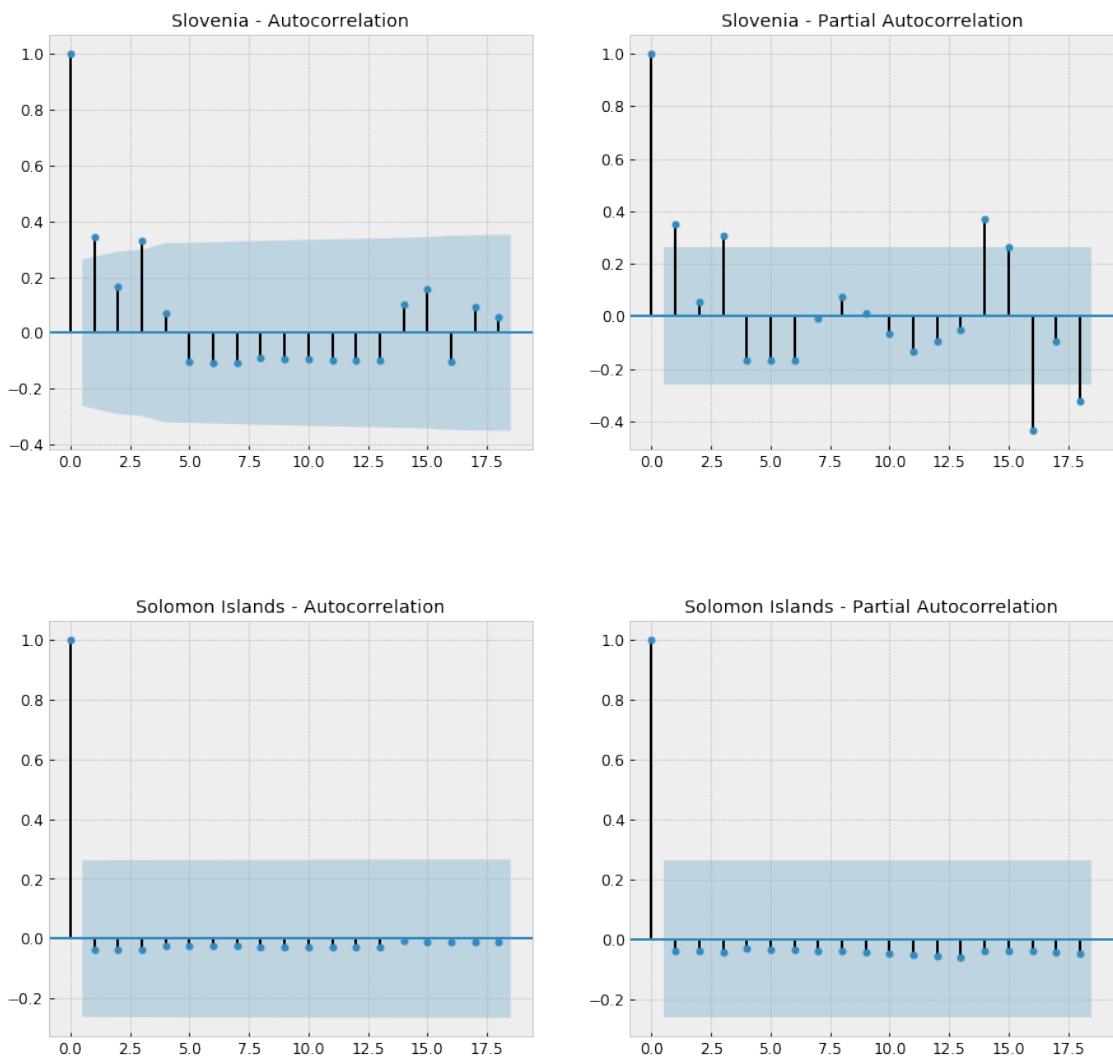


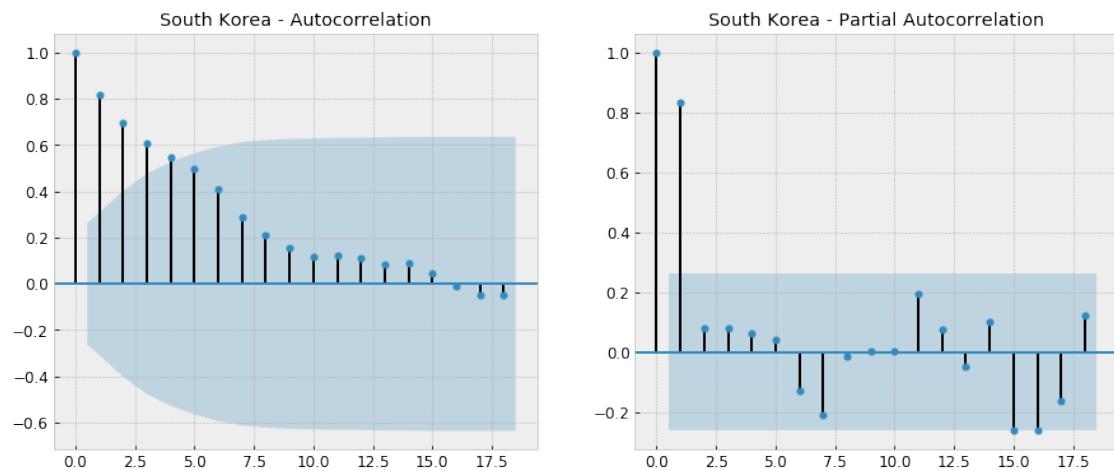
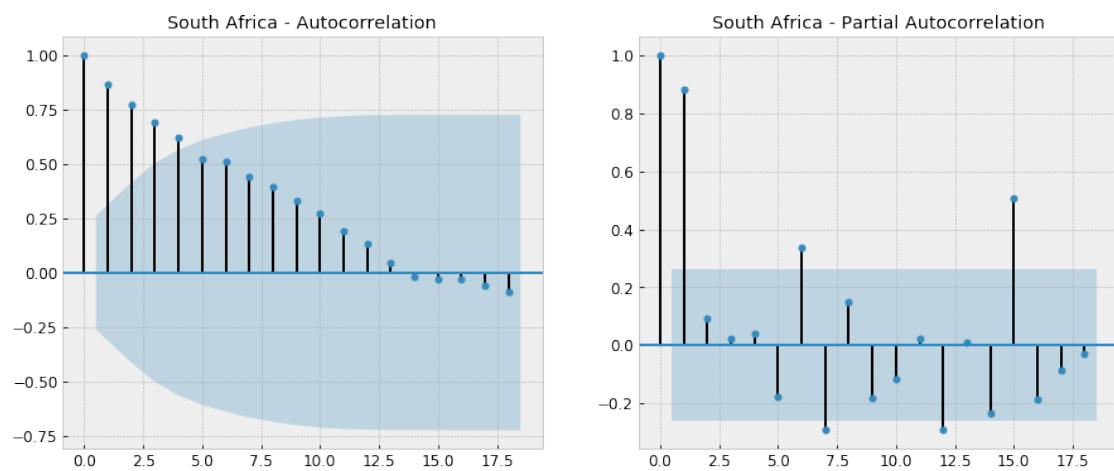
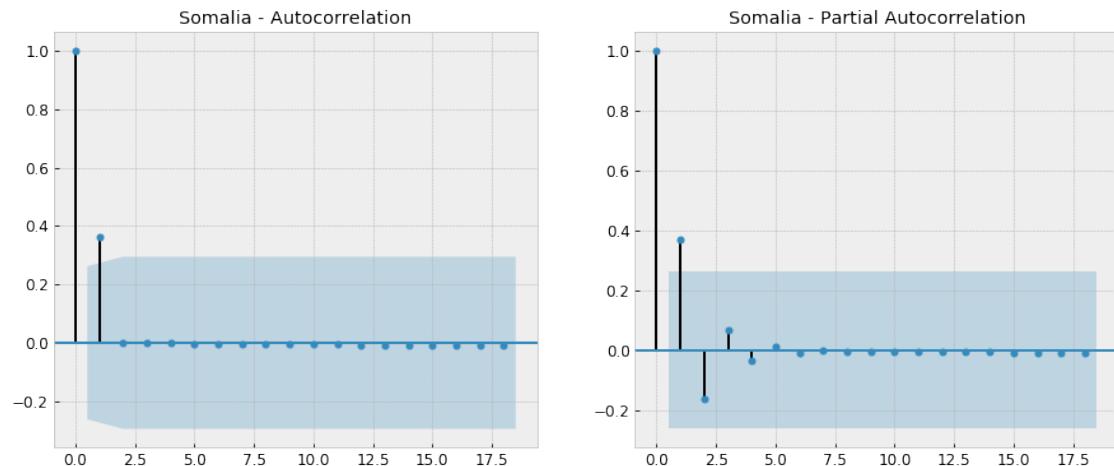


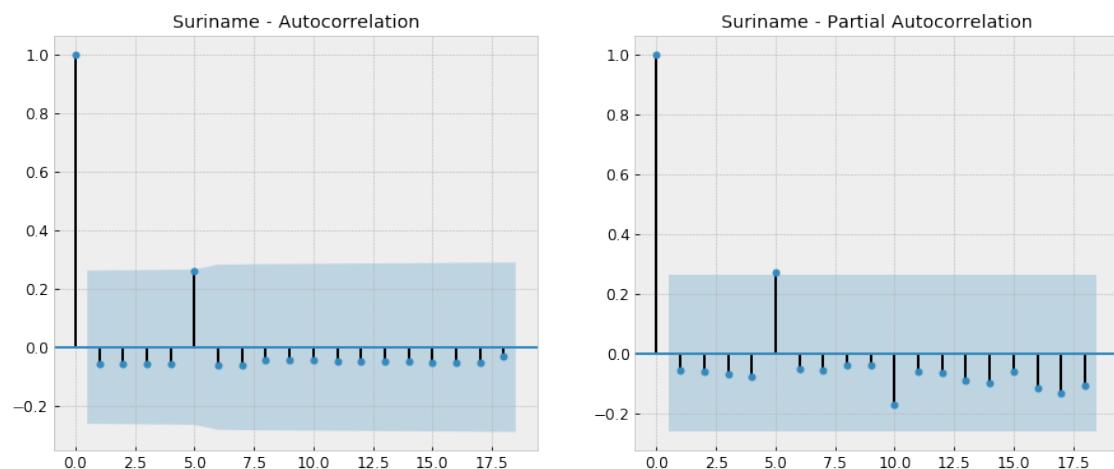
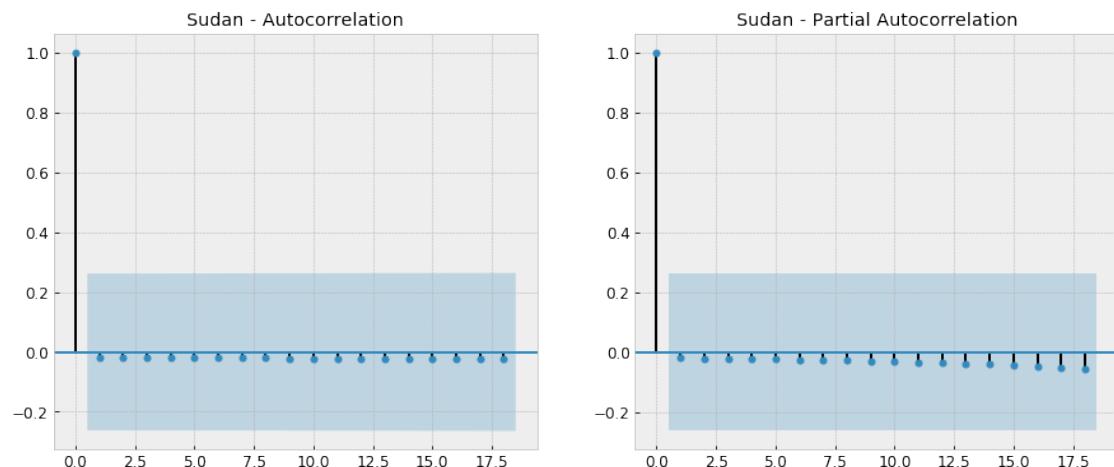
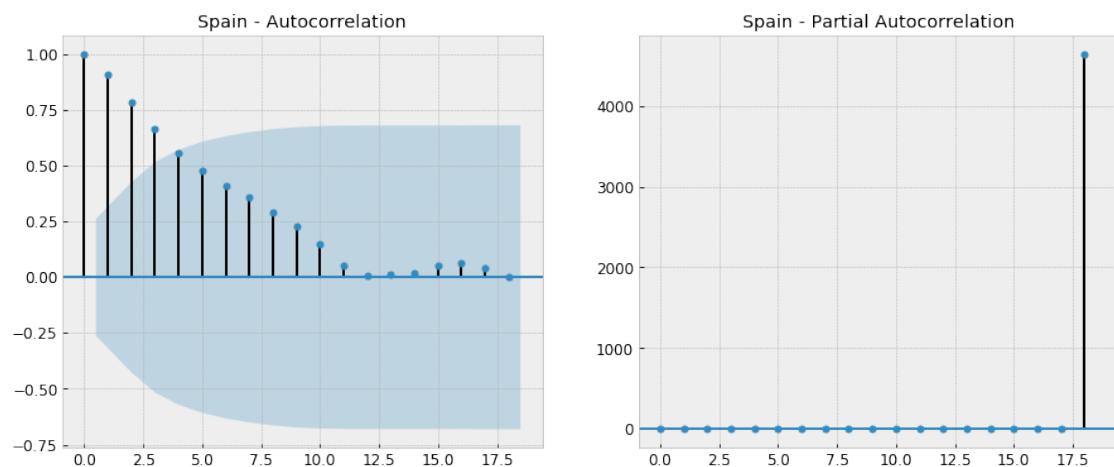


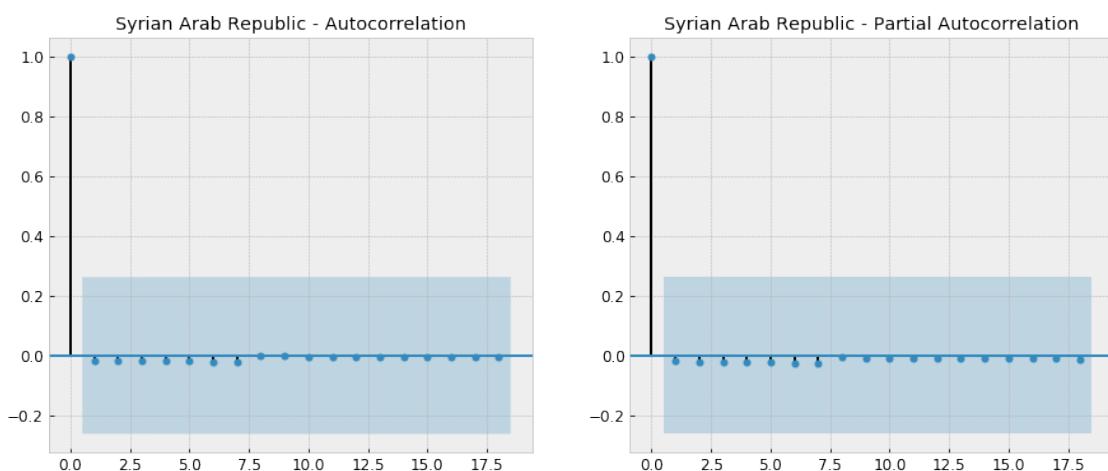
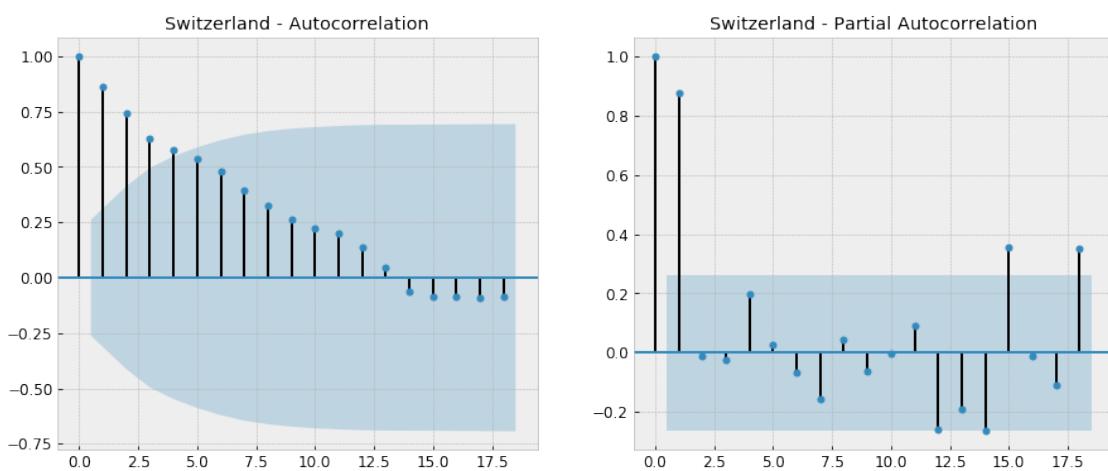
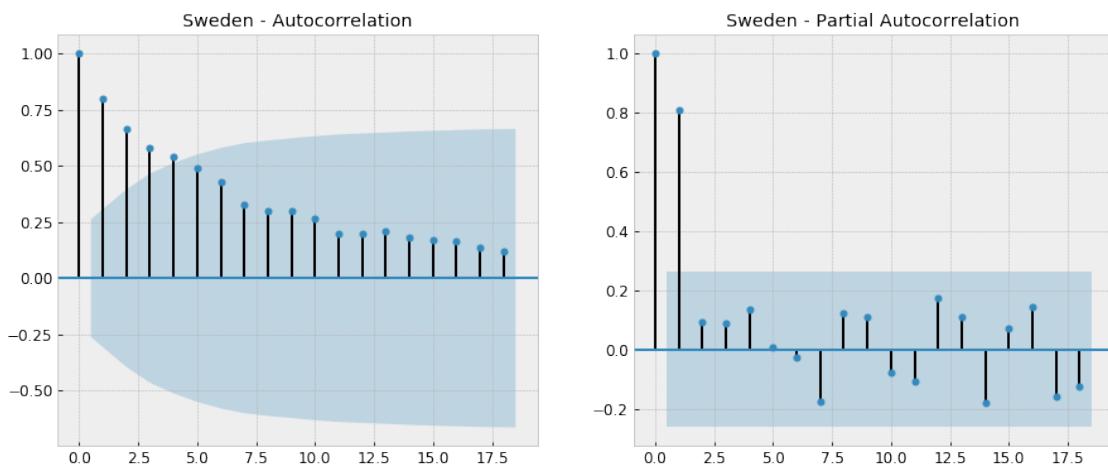


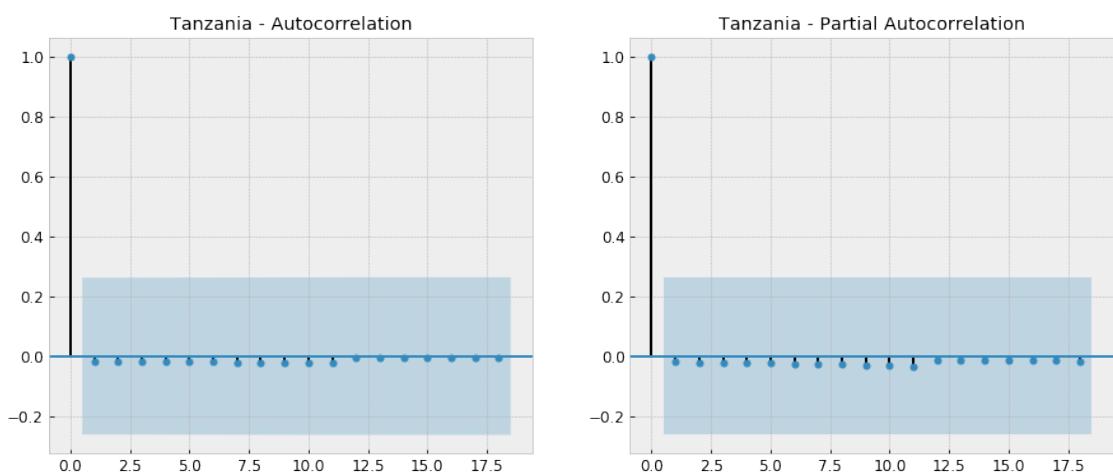
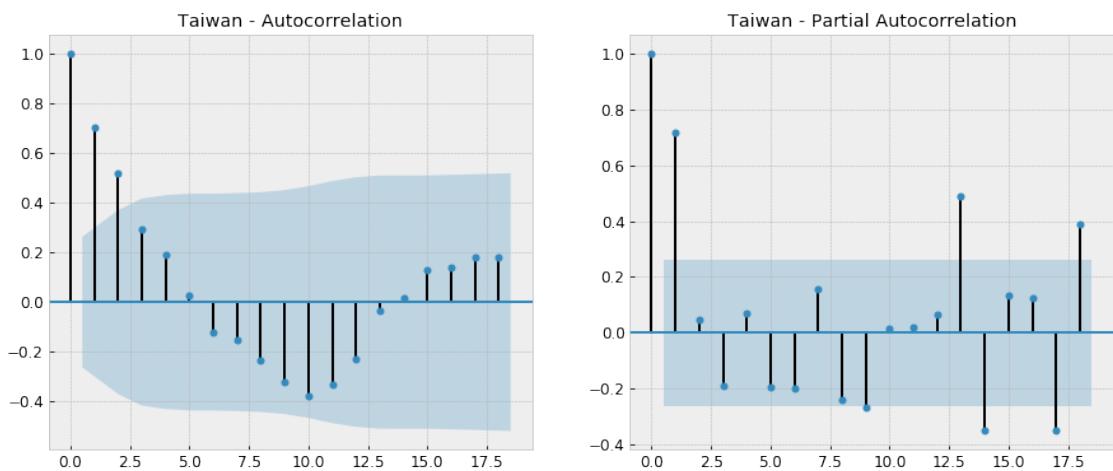


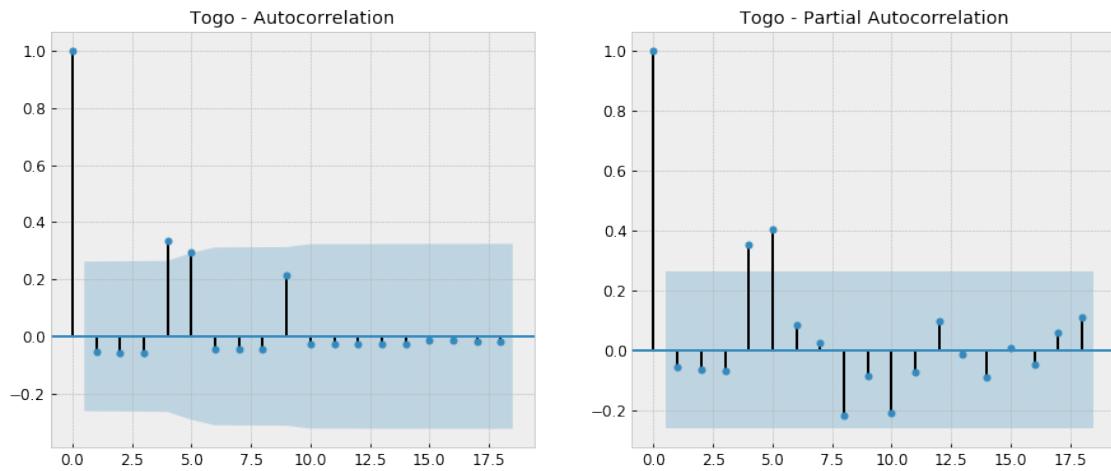
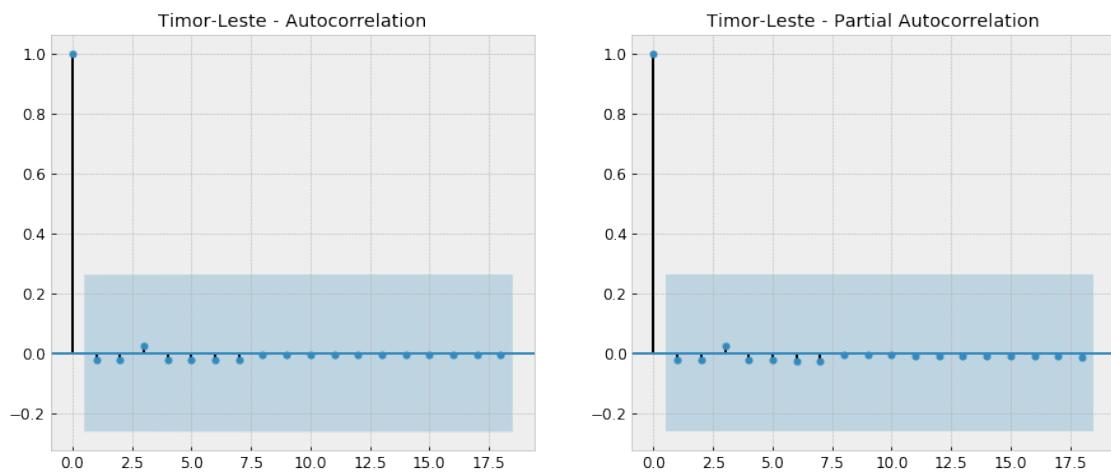
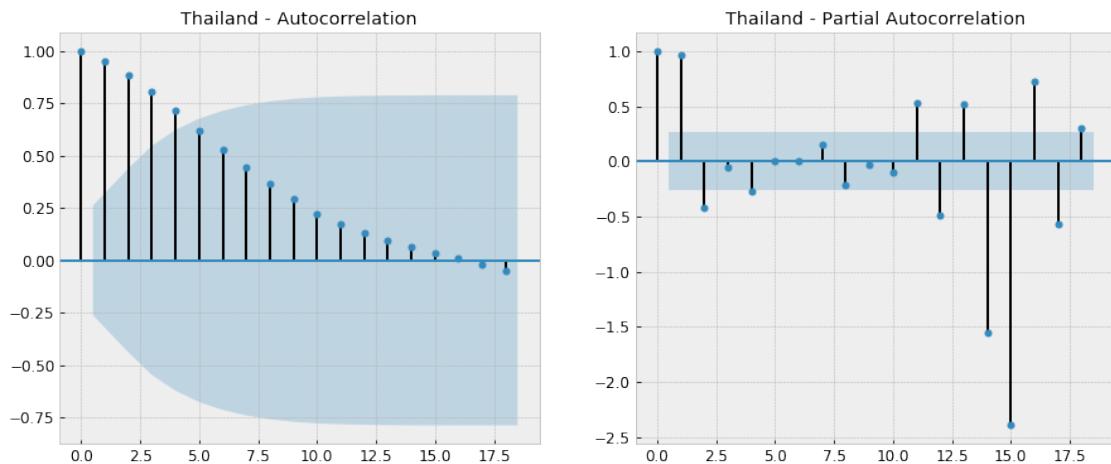


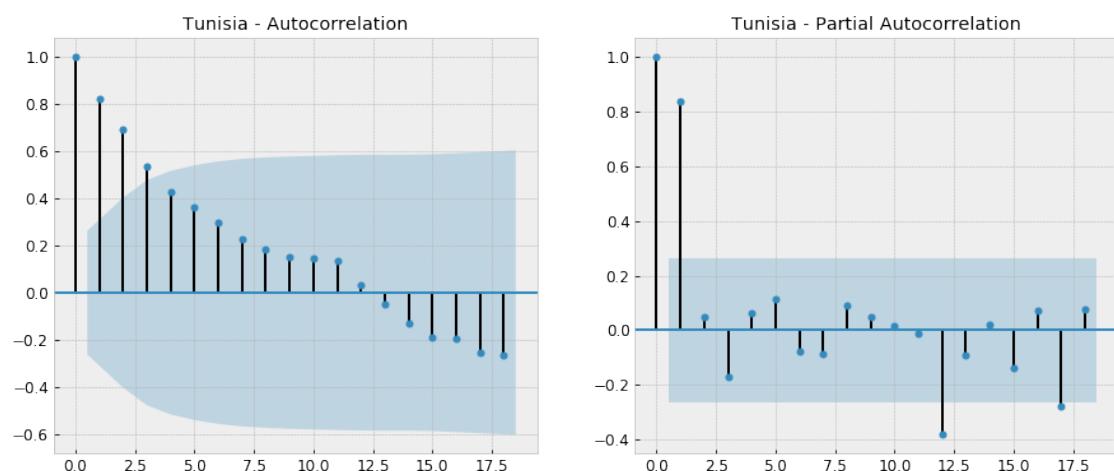
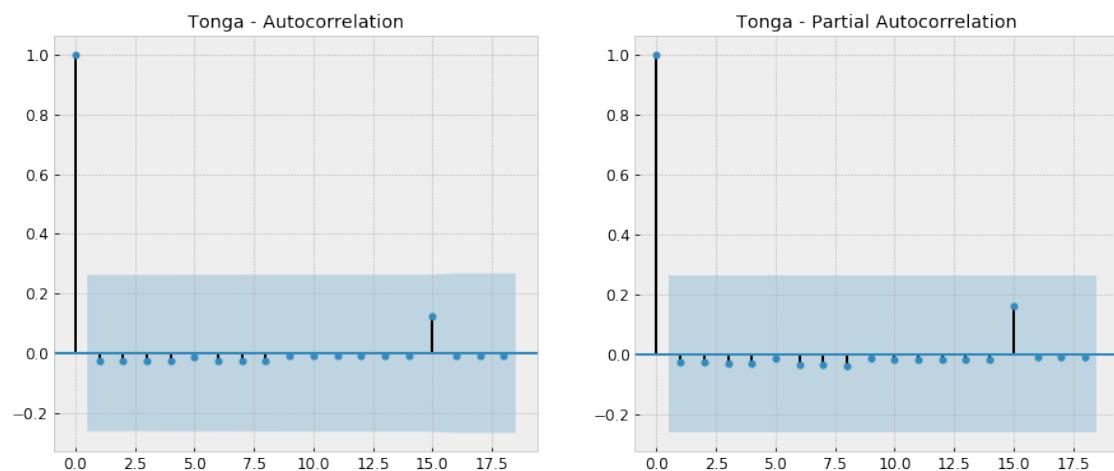
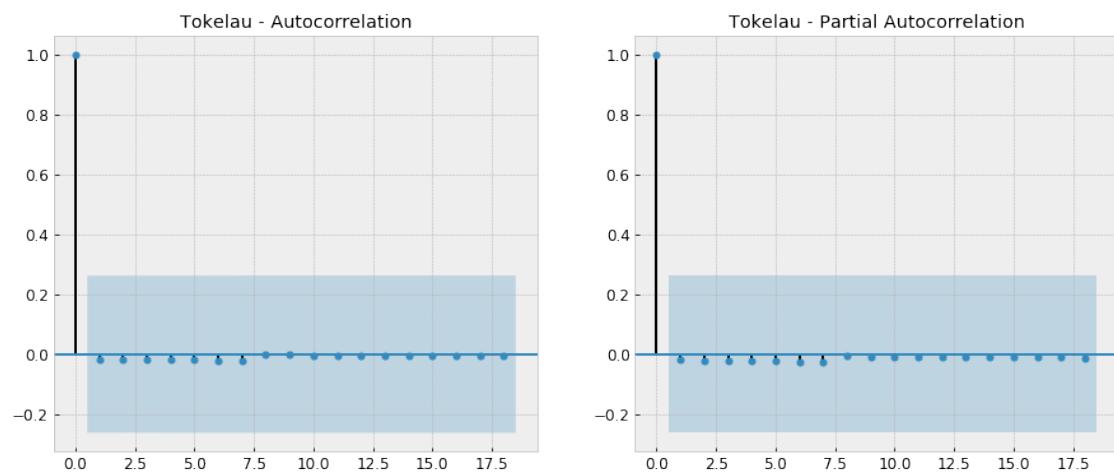


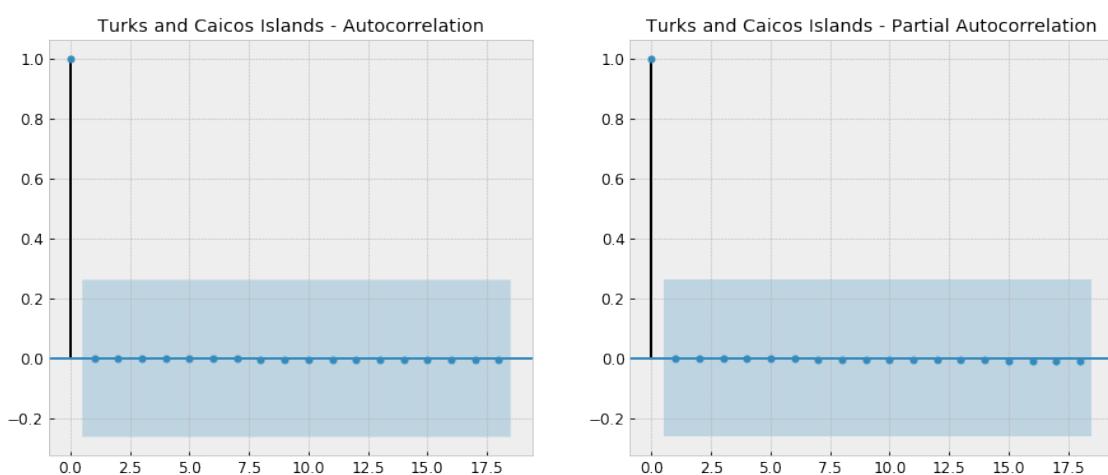
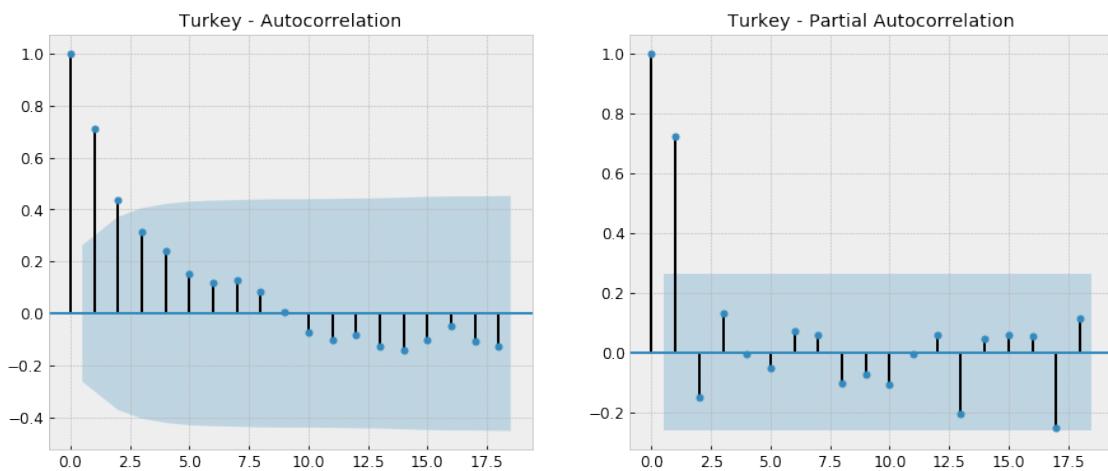


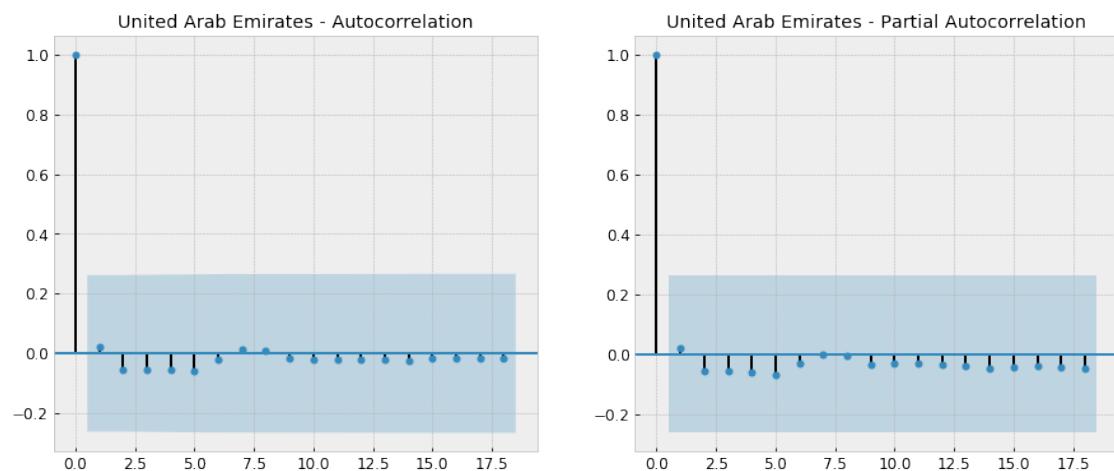
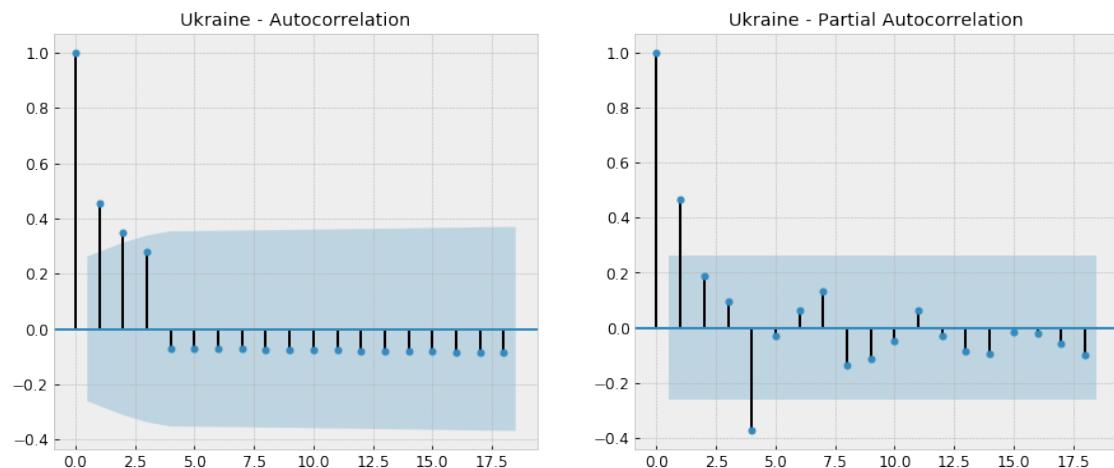
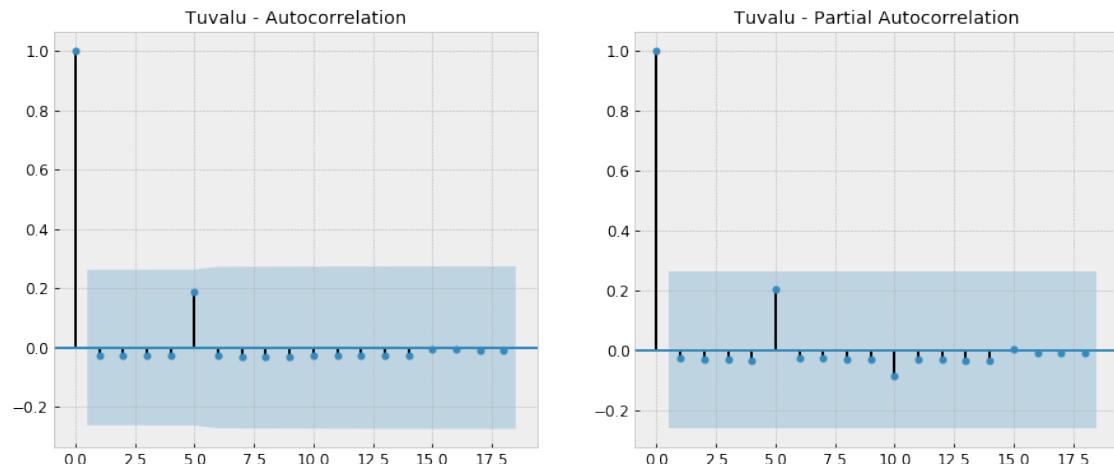


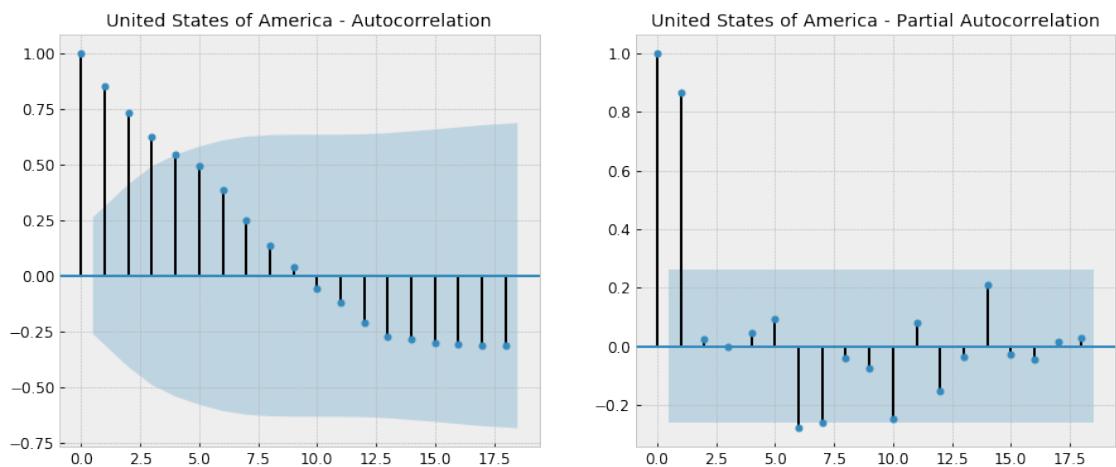
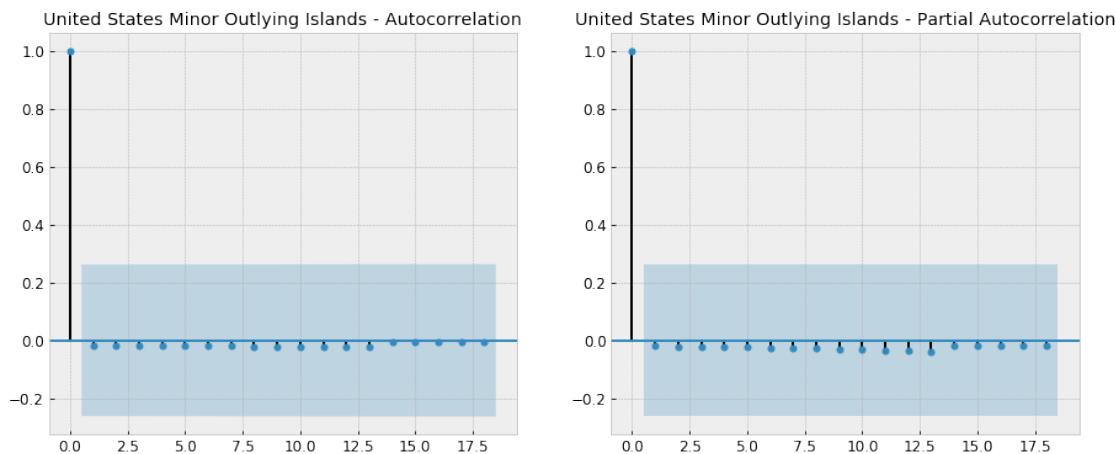
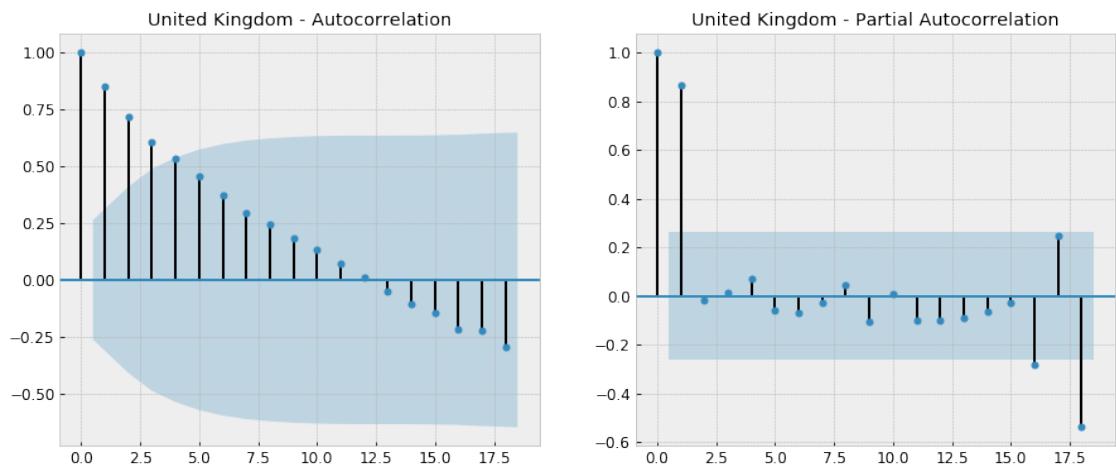


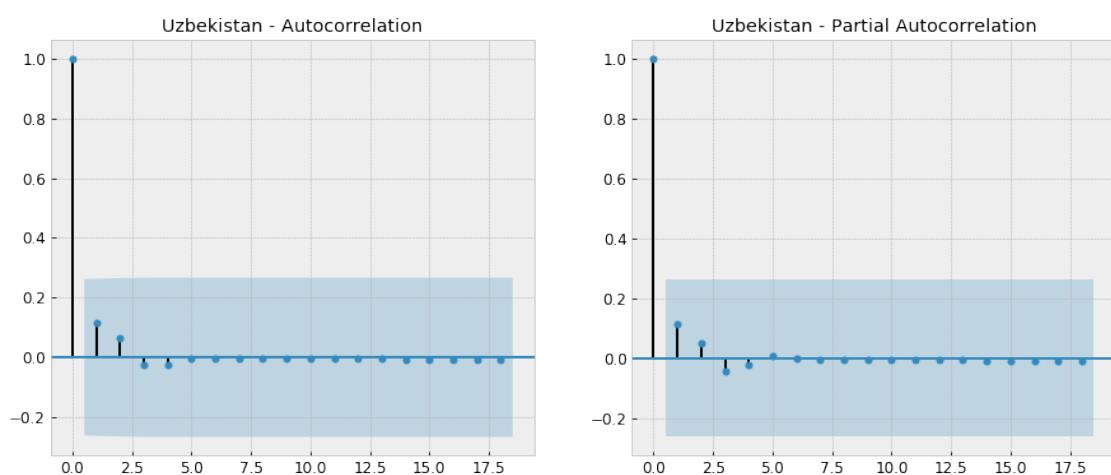
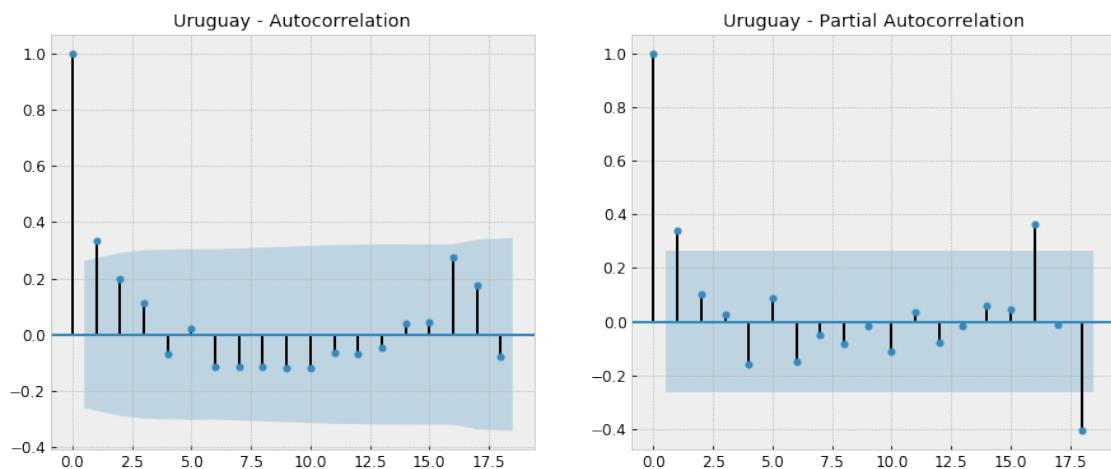


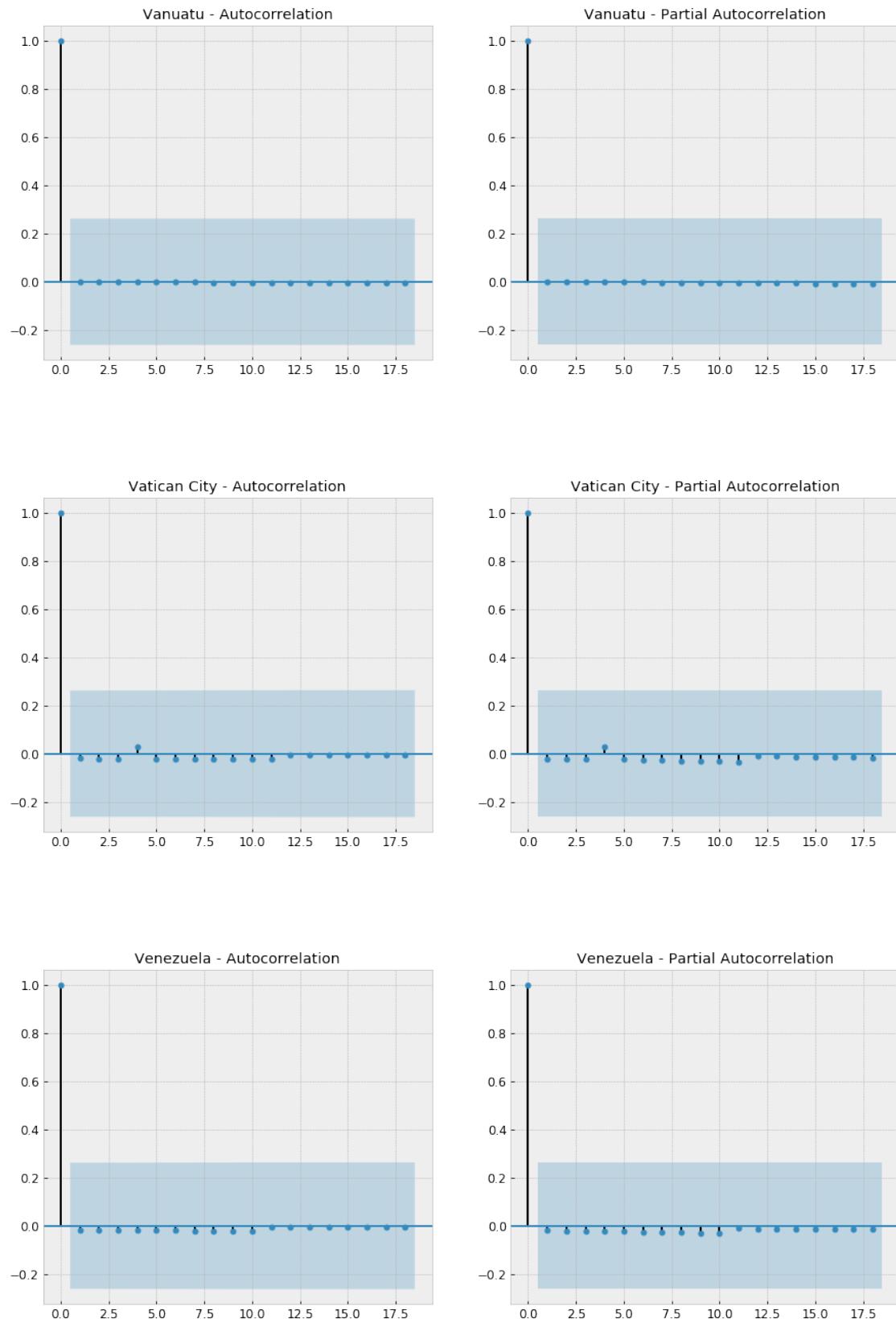


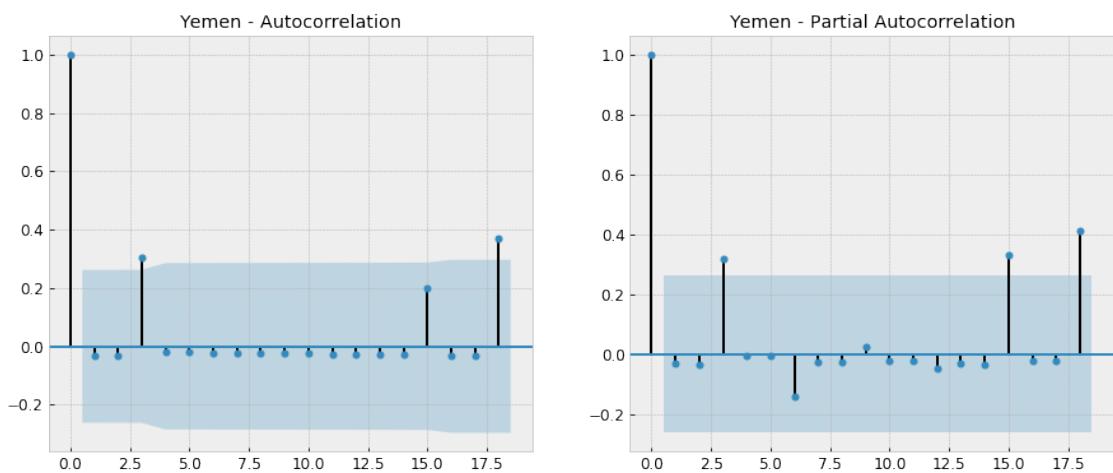
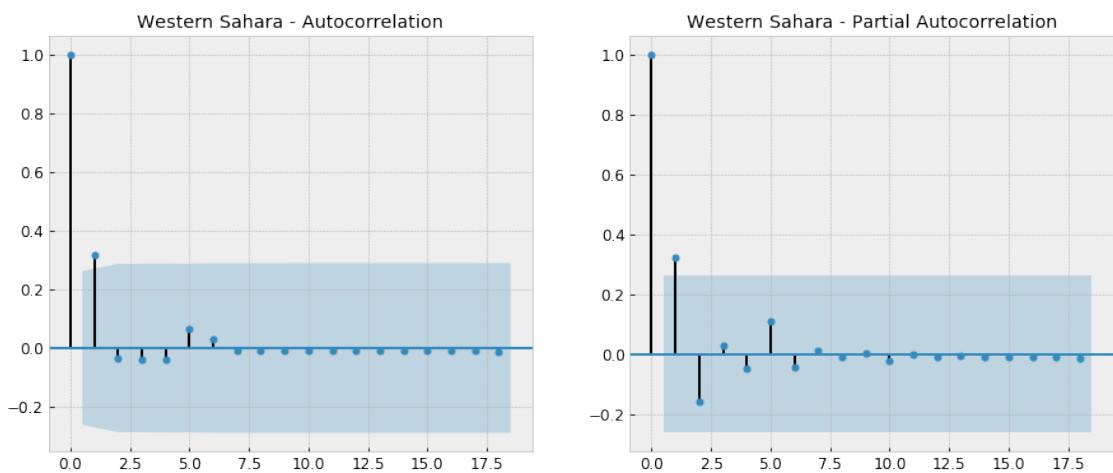
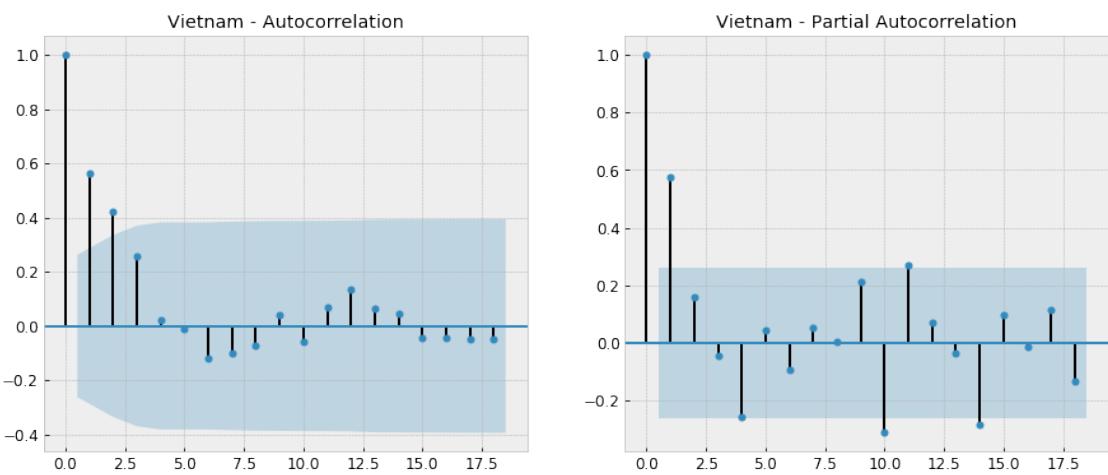


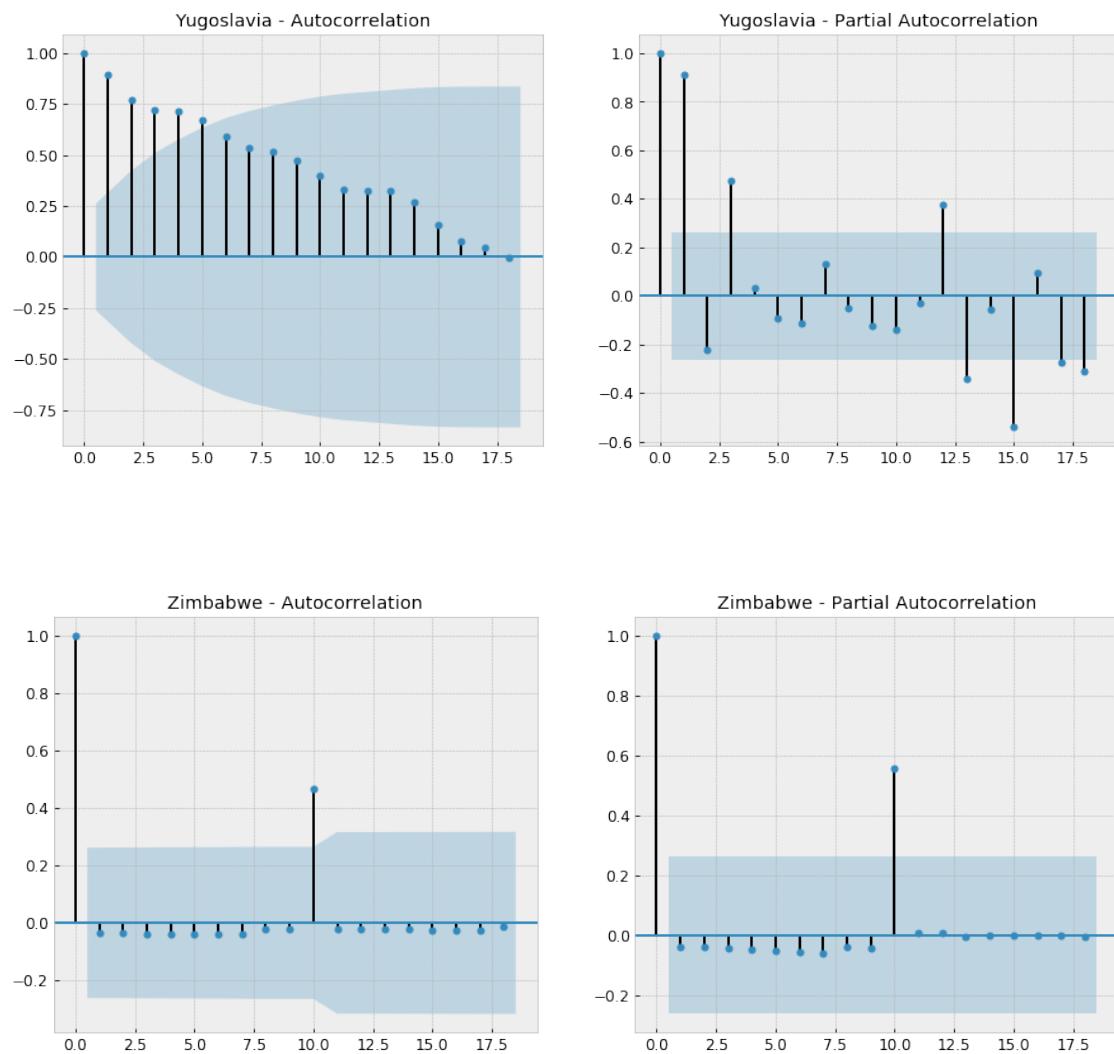


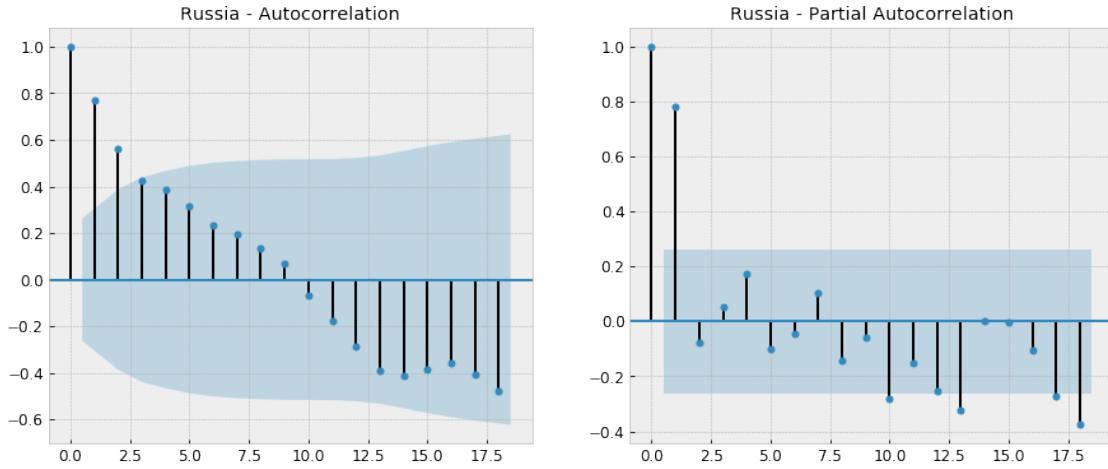












4 TODO

4.1 Interpreting ACF plots

| ACF Shape | Indicated Model | Countries fitting description |
|---|---|-------------------------------|
| Exponential, decaying to zero | Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model. | |
| Alternating positive and negative, decaying to zero | Use the partial autocorrelation plot to help identify the order. | |
| Autoregressive model. | | |
| One or more spikes, rest are essentially zero | Moving average model, order identified by where plot becomes zero. | |
| Decay, starting after a few lags | Mixed autoregressive and moving average (ARMA) model. | |
| All zero or close to zero | Data are essentially random. | |
| High values at fixed intervals | Include seasonal autoregressive term. | |
| No decay to zero | Series is not stationary | |

4.1.1 Plotting Rolling Statistics

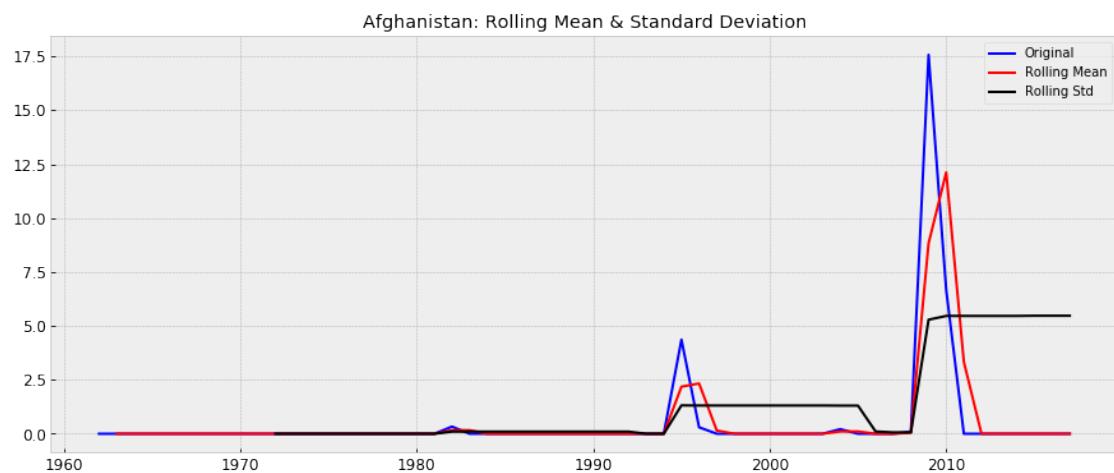
- The rolling mean and standard deviation are not constant with respect to time (increasing trend)
- The time series is hence not stationary

```
[178]: for country in countries:
    savefile = save_images(df_grps, 'rolling_stats')
    cycles = int((len(df_grps[country]))/5)
    plt.figure(figsize=(15,6))

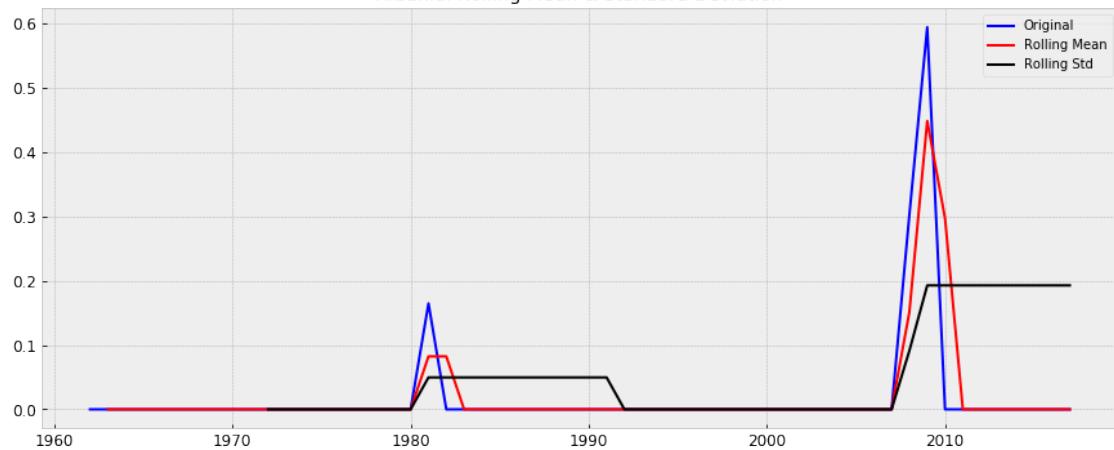
    rolmean = df_grps[country].rolling(2).mean()
    rolstd = df_grps[country].rolling(cycles).std()

    #Plot rolling statistics:
    orig = plt.plot(df_grps[country], color='blue',label='Original')
    mean = plt.plot(rolmean, color='red', label='Rolling Mean')
    std = plt.plot(rolstd, color='black', label = 'Rolling Std')

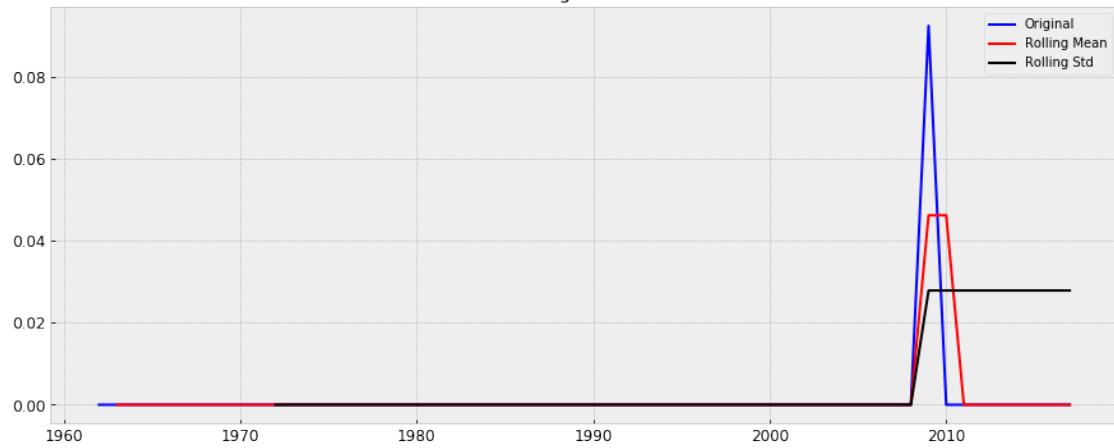
    plt.legend(loc='best')
    plt.title('{}: Rolling Mean & Standard Deviation'.format(df_grps[country].
    ↪name))
    plt.savefig(str('{}_Rolling_Stats'.format(savefile)))
    plt.show(block=False)
```



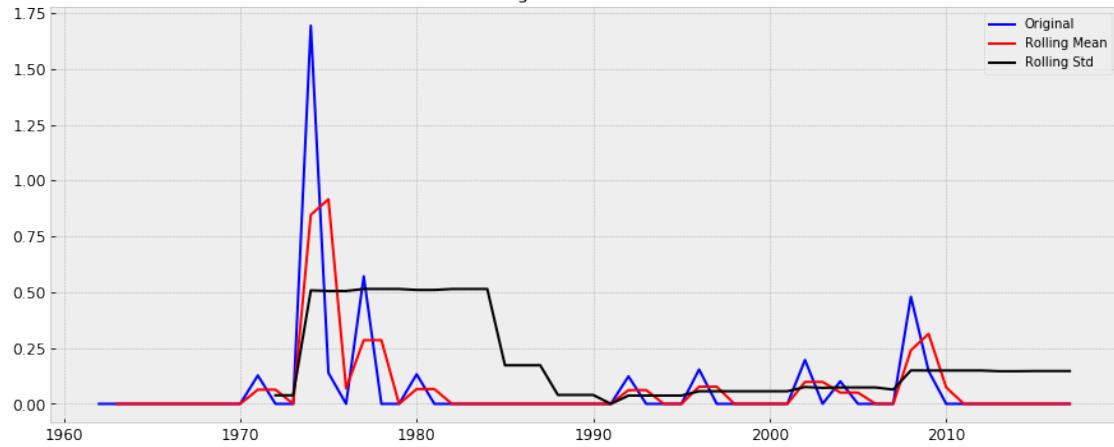
Albania: Rolling Mean & Standard Deviation



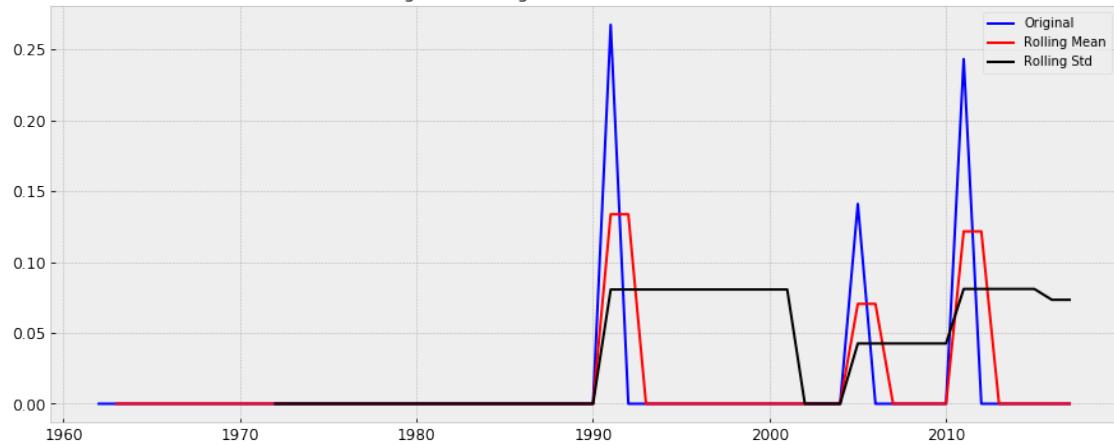
American Samoa: Rolling Mean & Standard Deviation



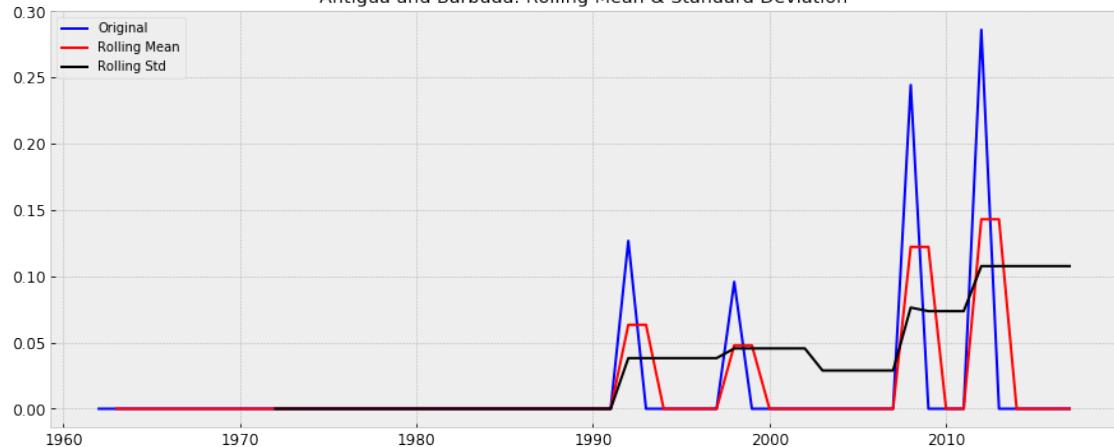
Andorra: Rolling Mean & Standard Deviation



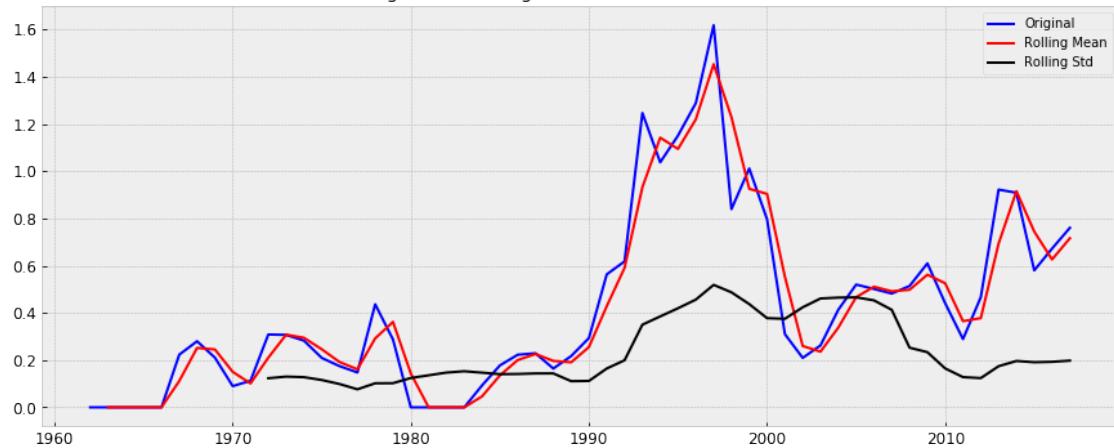
Anguilla: Rolling Mean & Standard Deviation



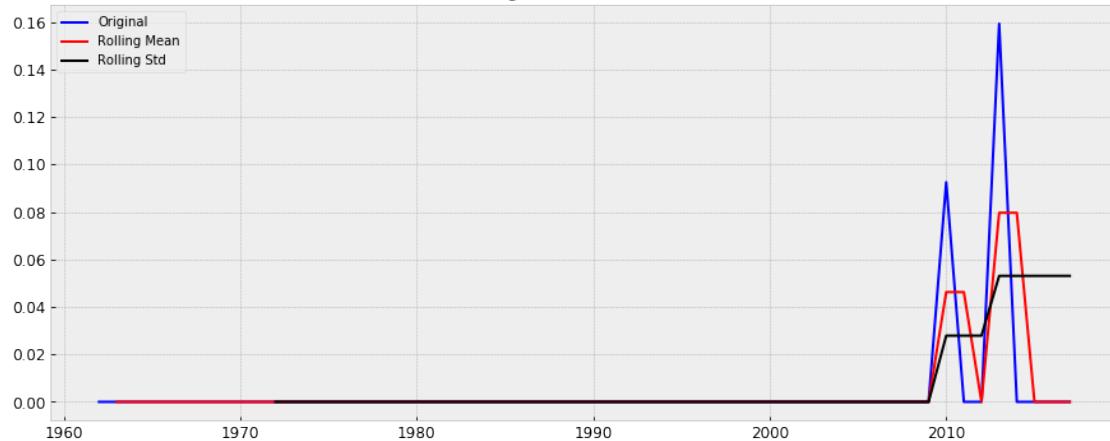
Antigua and Barbuda: Rolling Mean & Standard Deviation



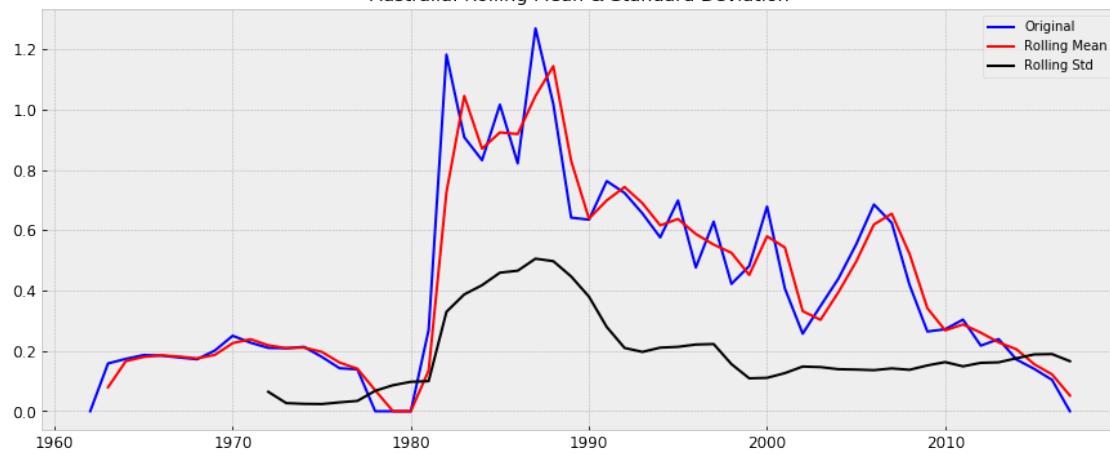
Argentina: Rolling Mean & Standard Deviation



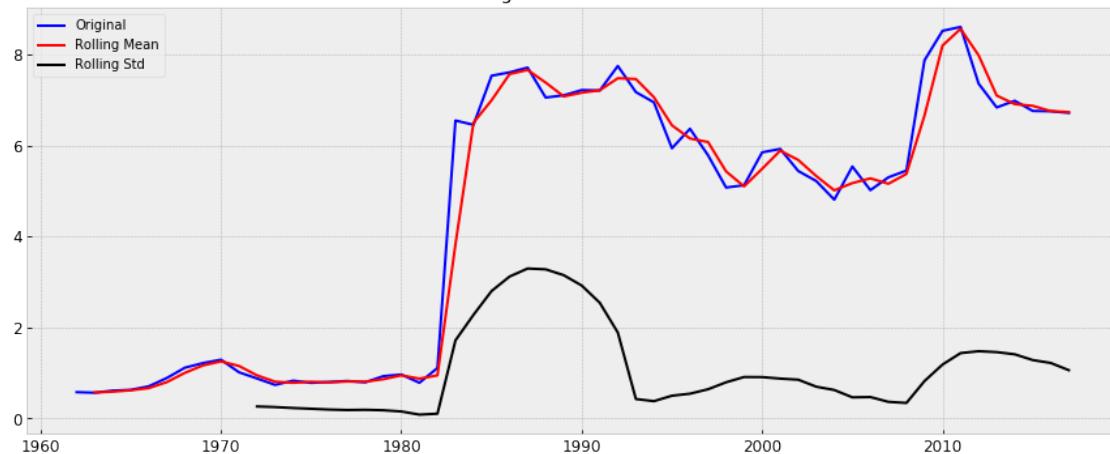
Aruba: Rolling Mean & Standard Deviation



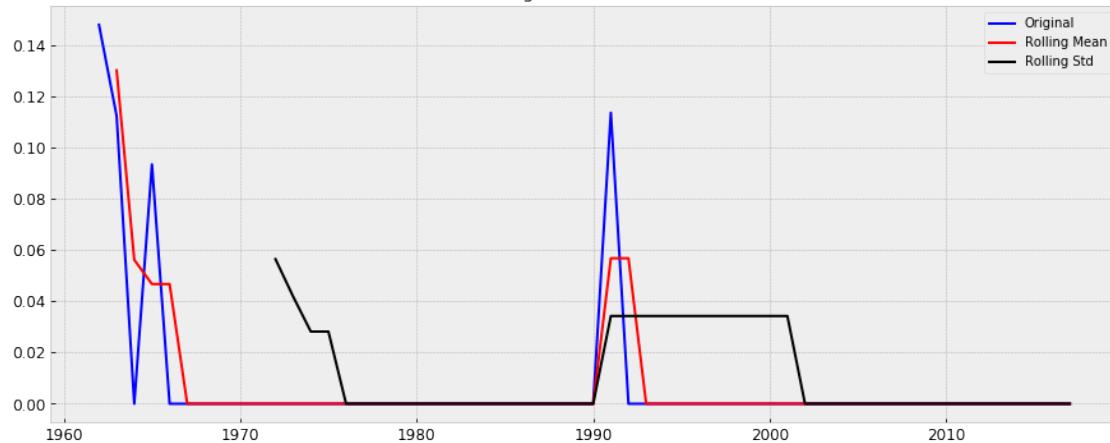
Australia: Rolling Mean & Standard Deviation



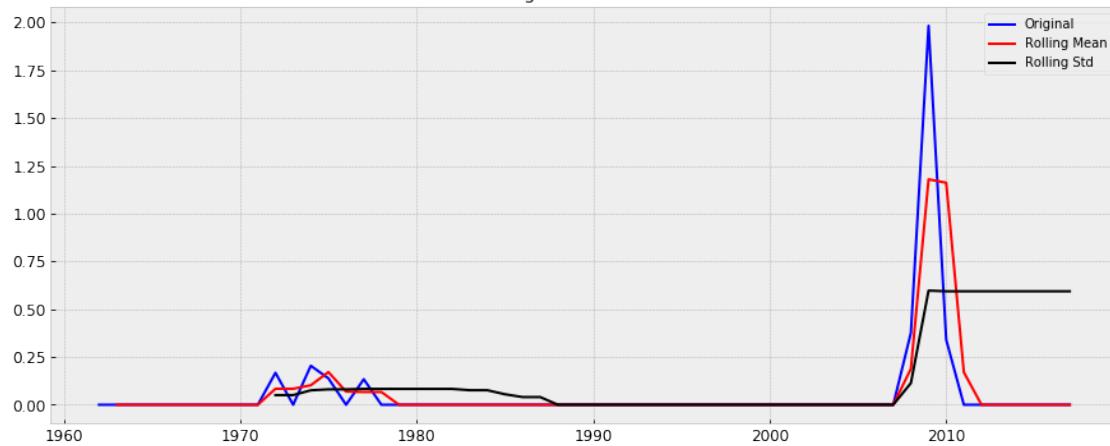
Austria: Rolling Mean & Standard Deviation



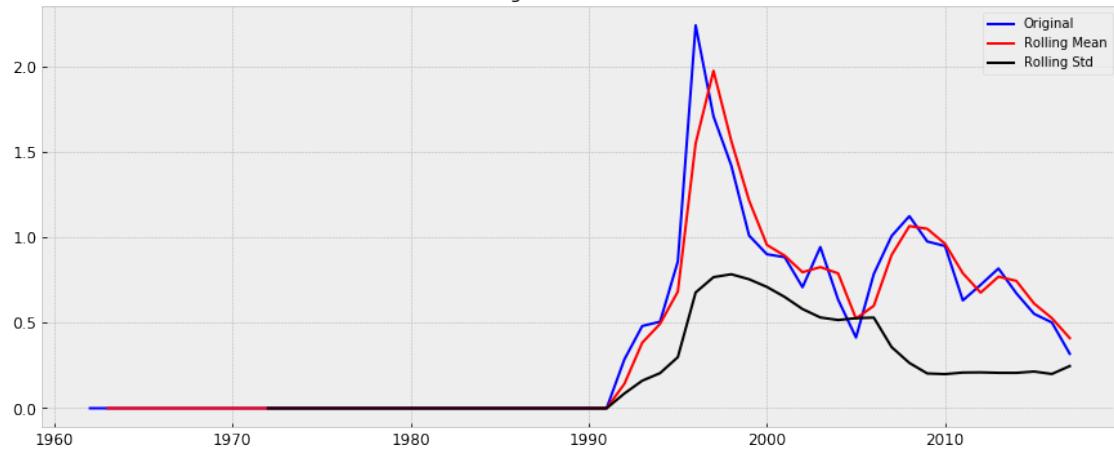
Bahamas: Rolling Mean & Standard Deviation



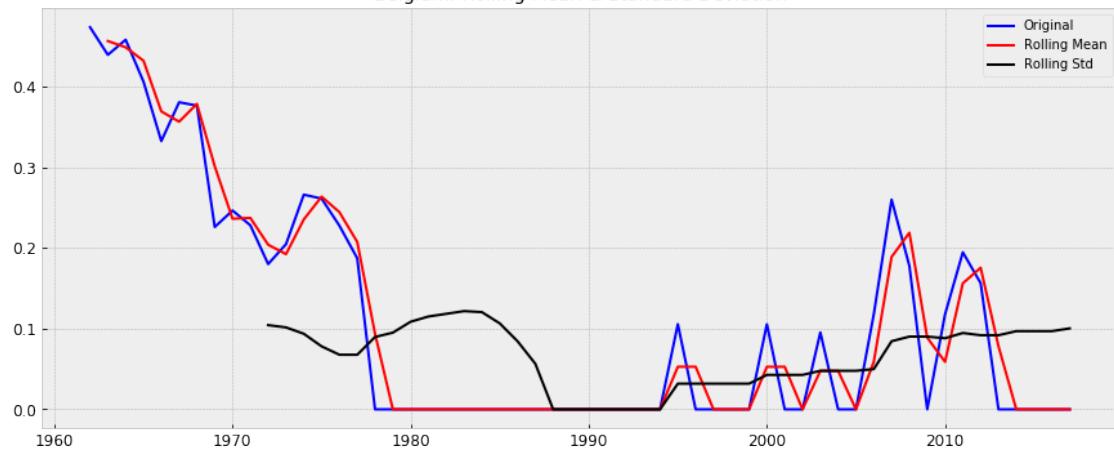
Barbados: Rolling Mean & Standard Deviation



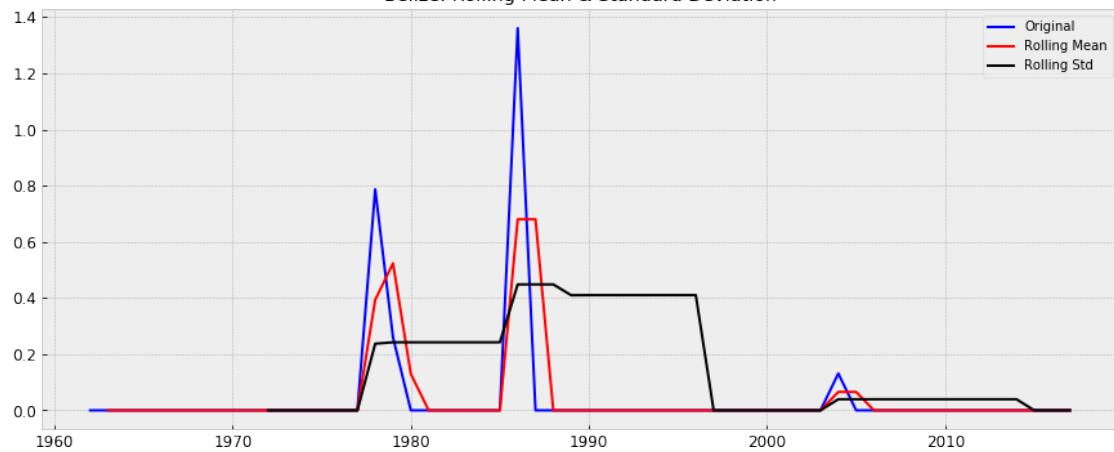
Belarus: Rolling Mean & Standard Deviation



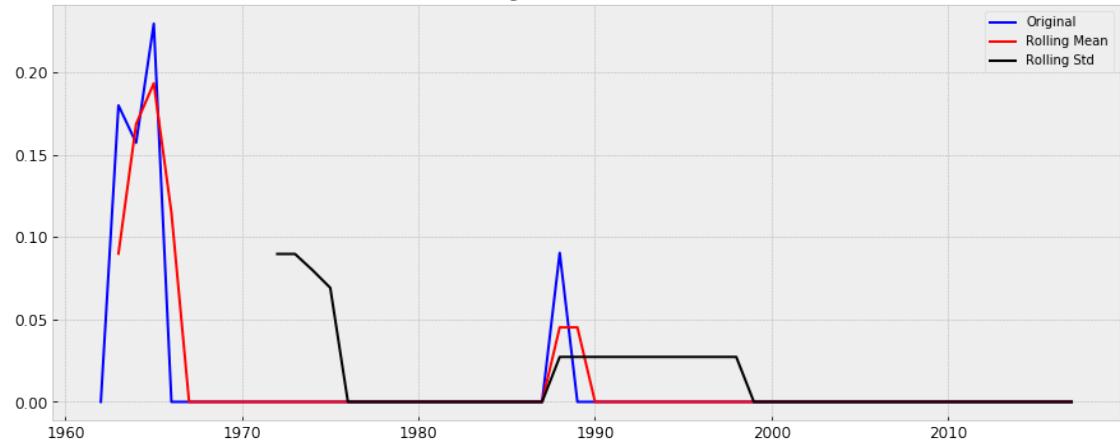
Belgium: Rolling Mean & Standard Deviation



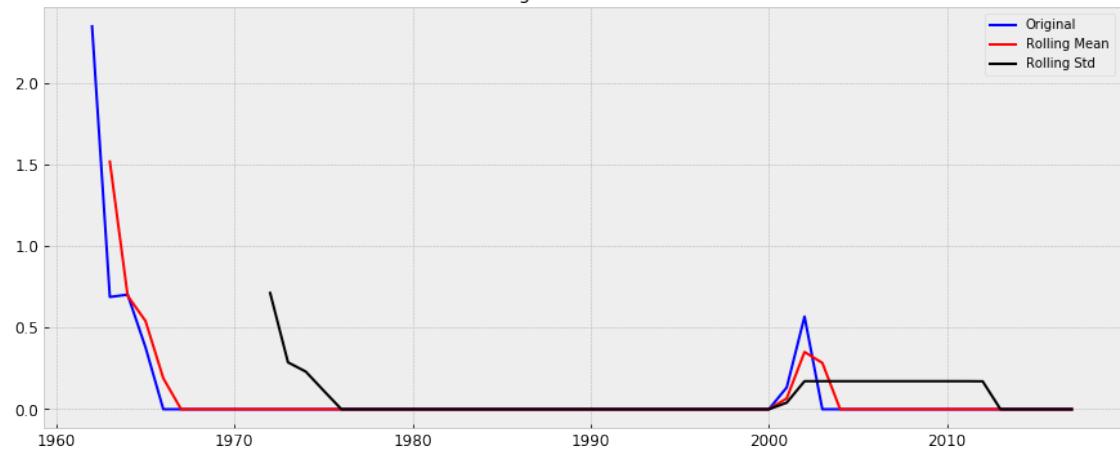
Belize: Rolling Mean & Standard Deviation



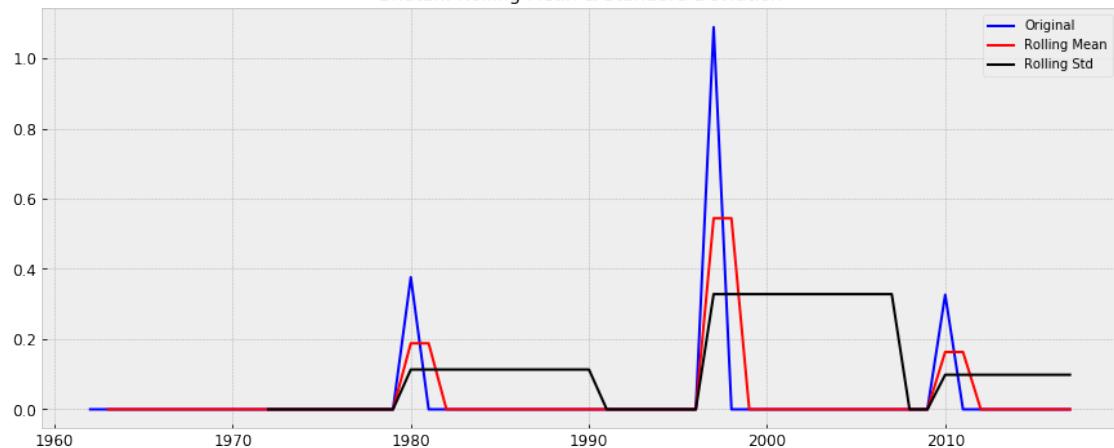
Benin: Rolling Mean & Standard Deviation



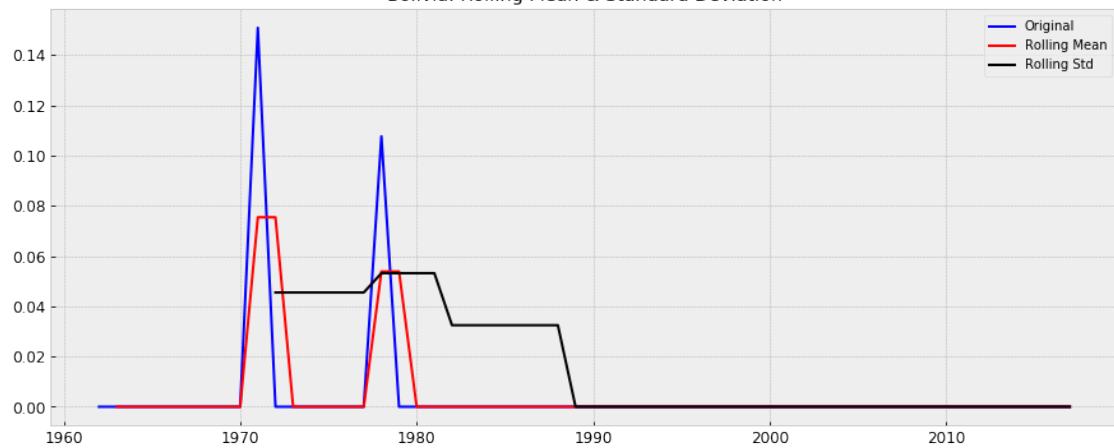
Bermuda: Rolling Mean & Standard Deviation



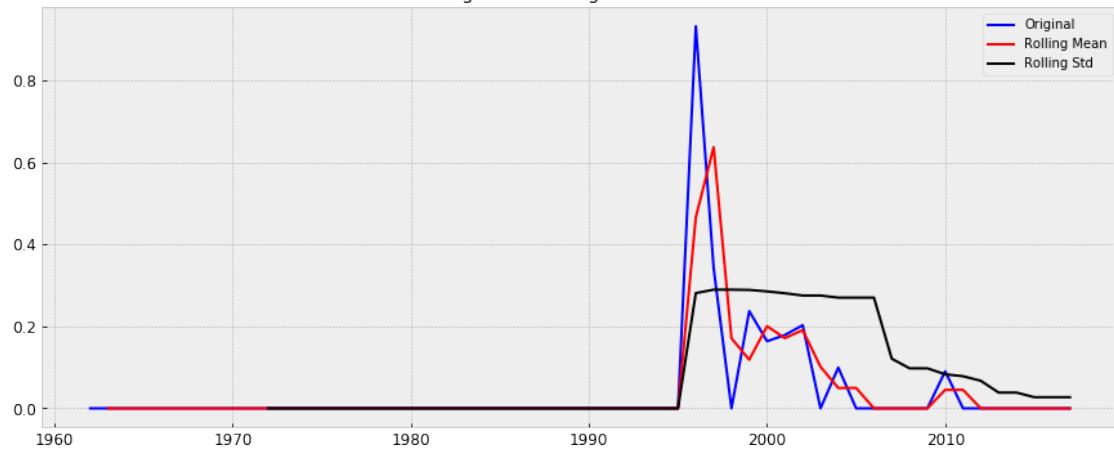
Bhutan: Rolling Mean & Standard Deviation

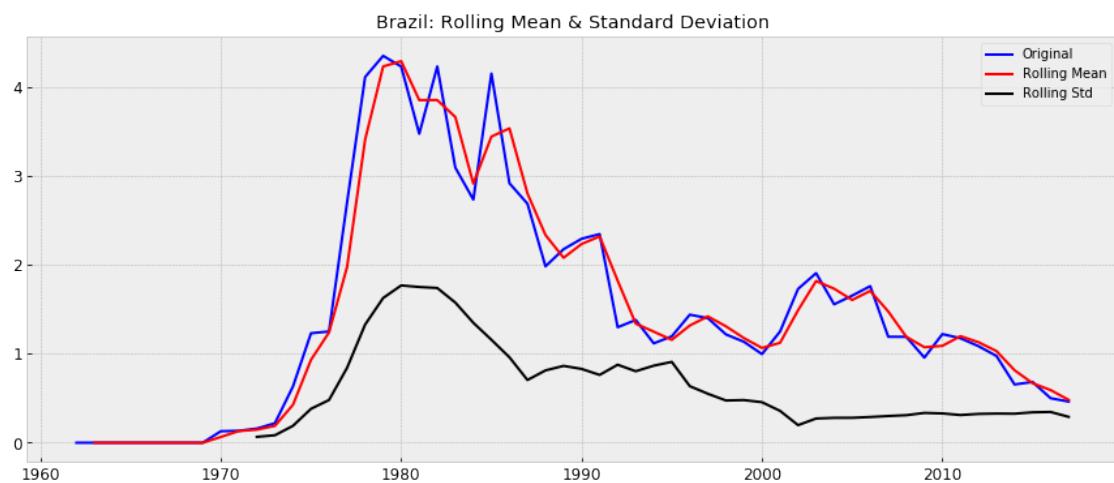
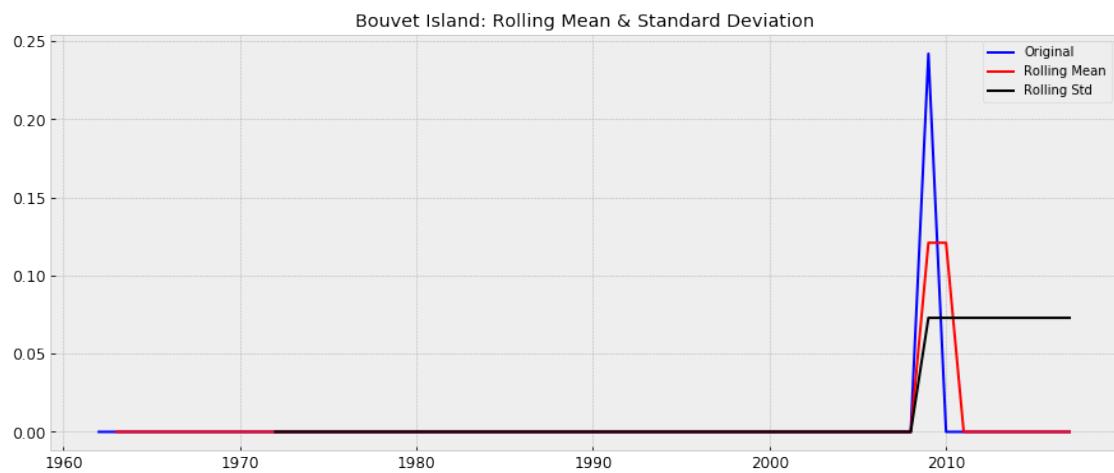
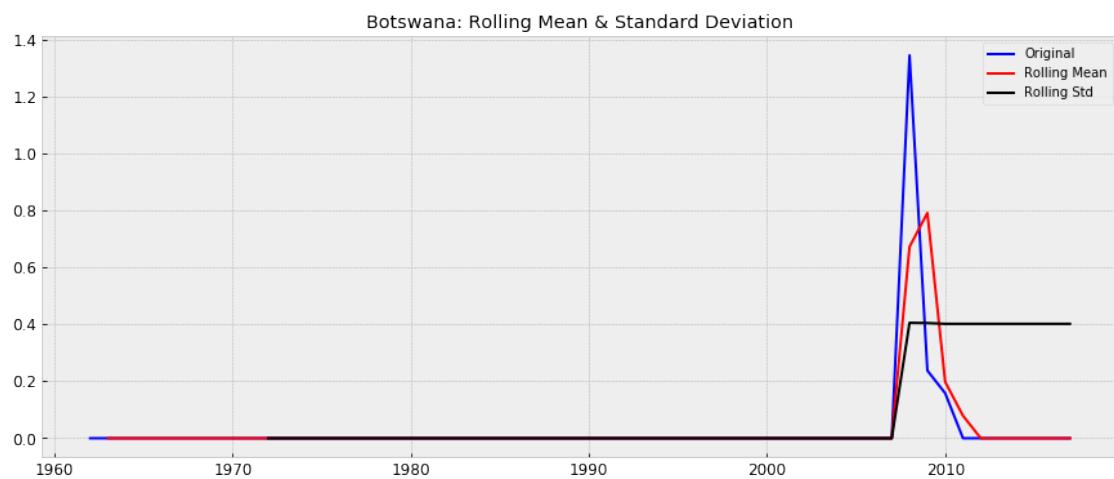


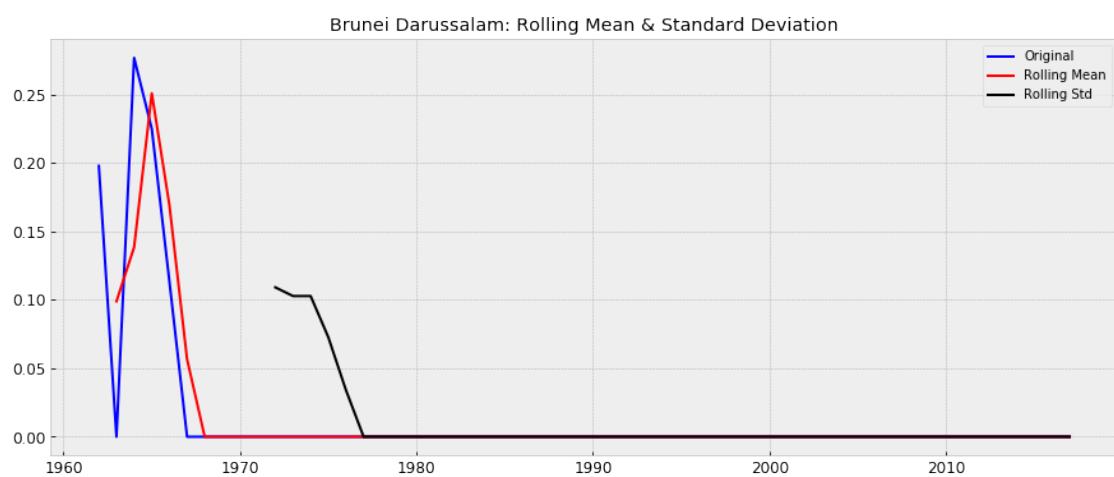
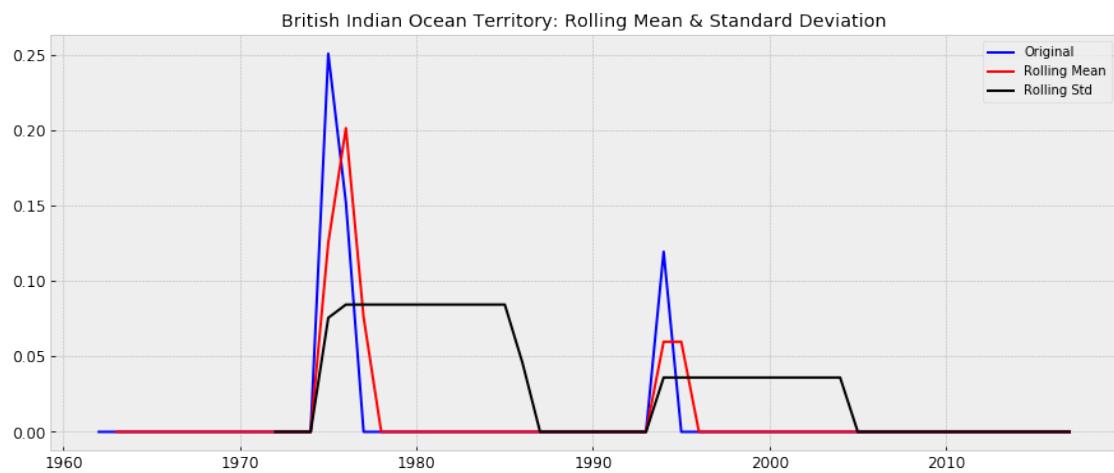
Bolivia: Rolling Mean & Standard Deviation



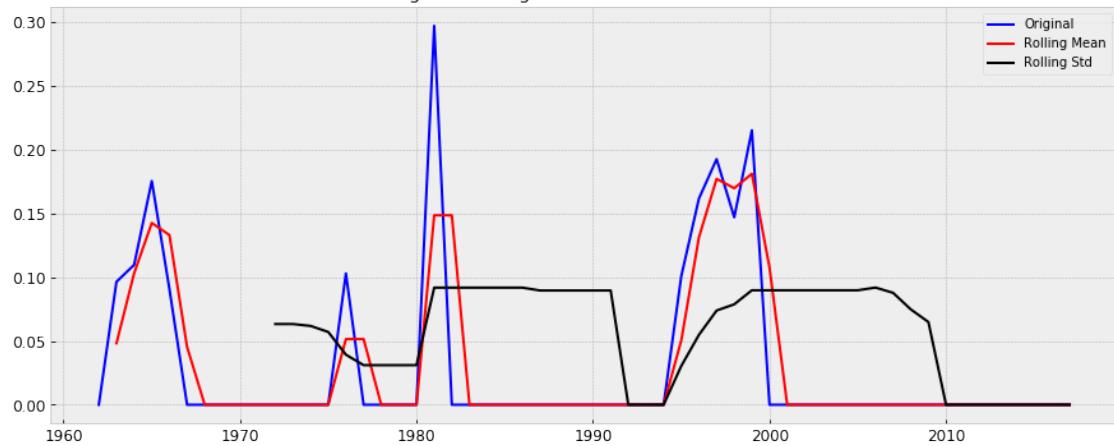
Bosnia and Herzegovina: Rolling Mean & Standard Deviation



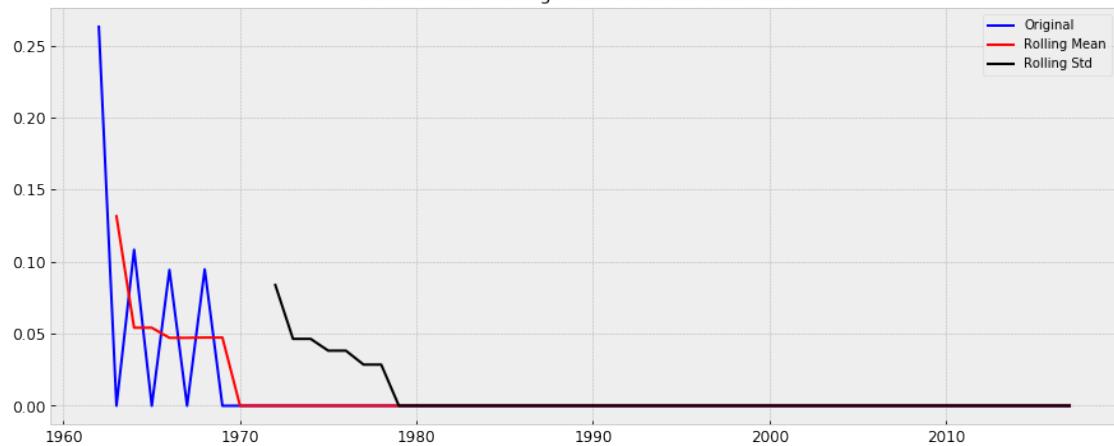




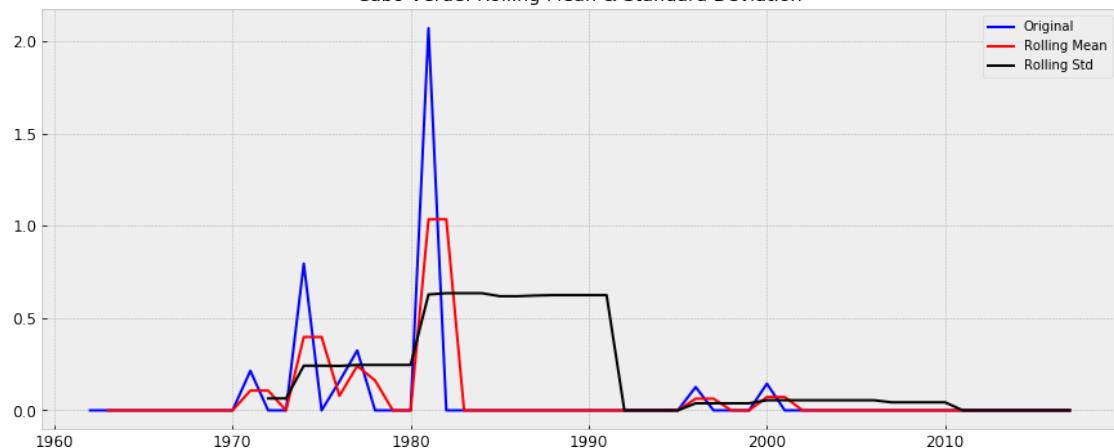
Bulgaria: Rolling Mean & Standard Deviation

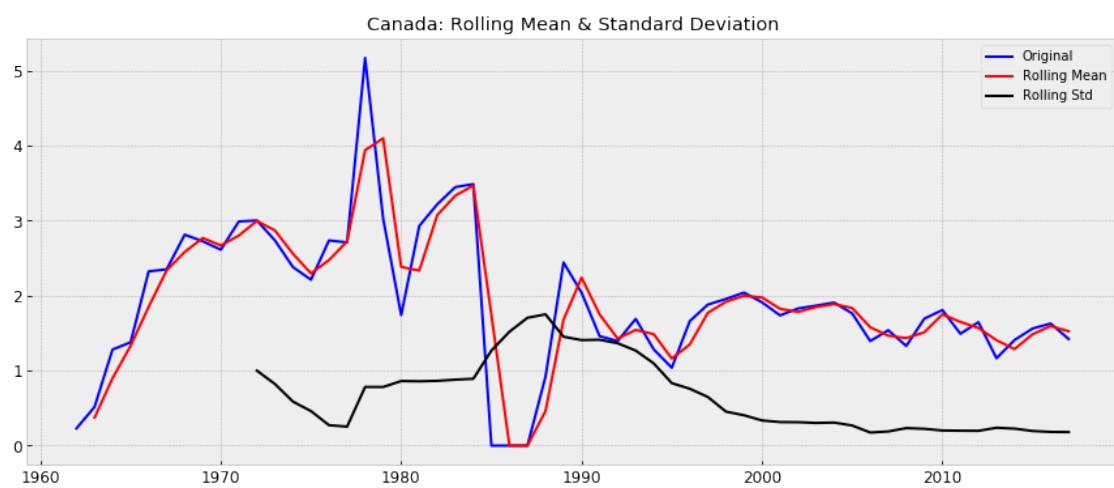
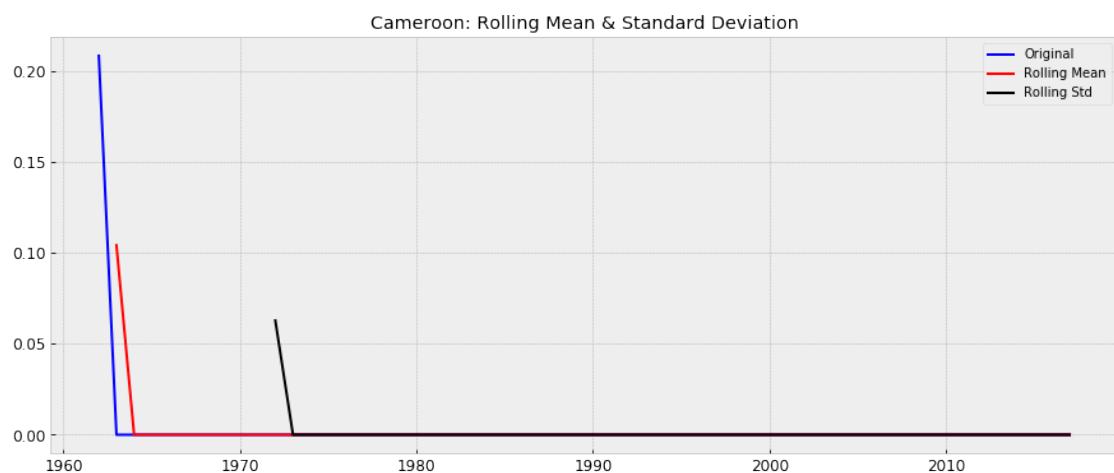
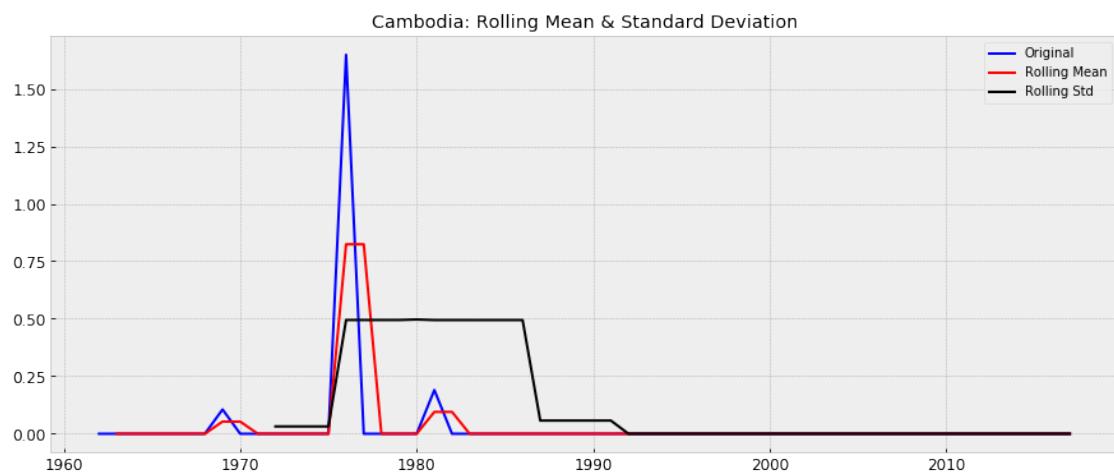


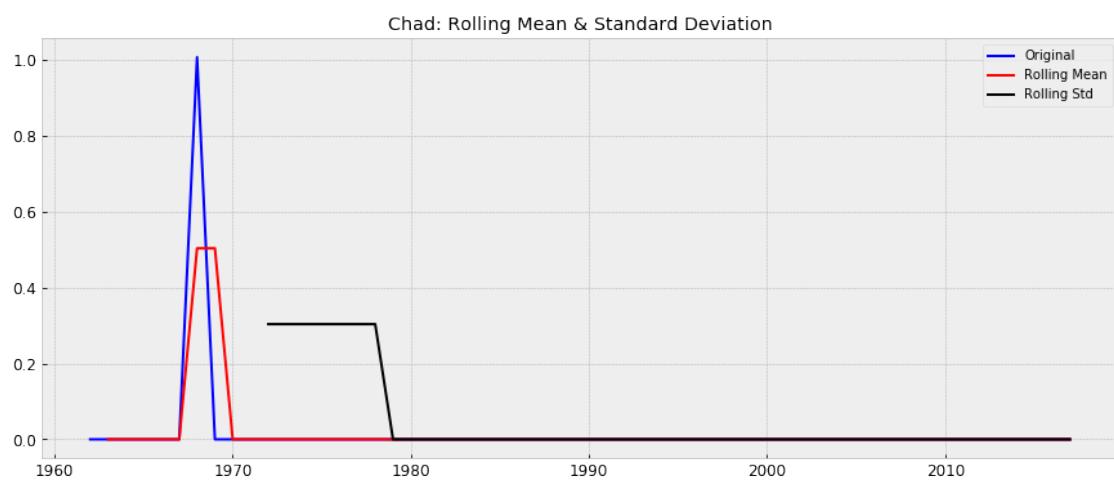
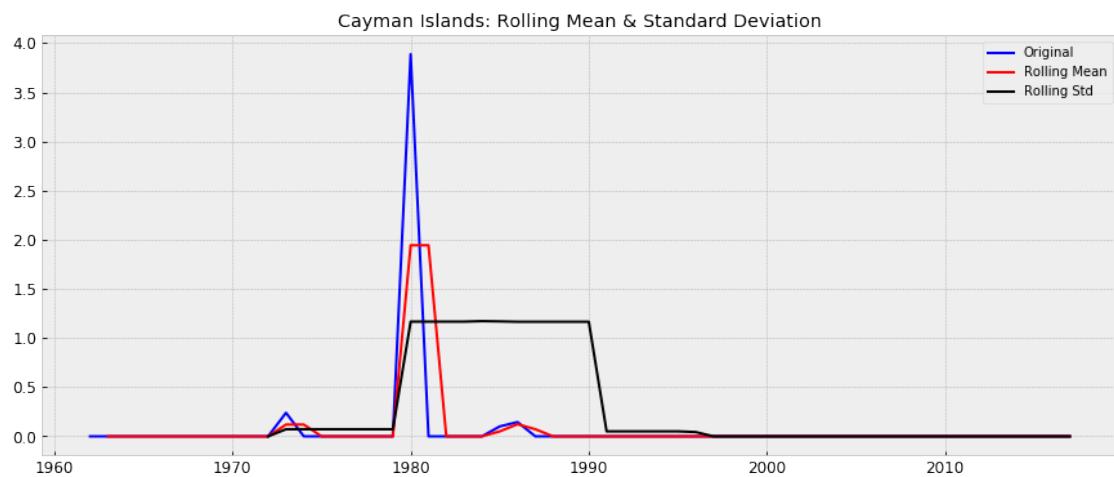
Burkina Faso: Rolling Mean & Standard Deviation



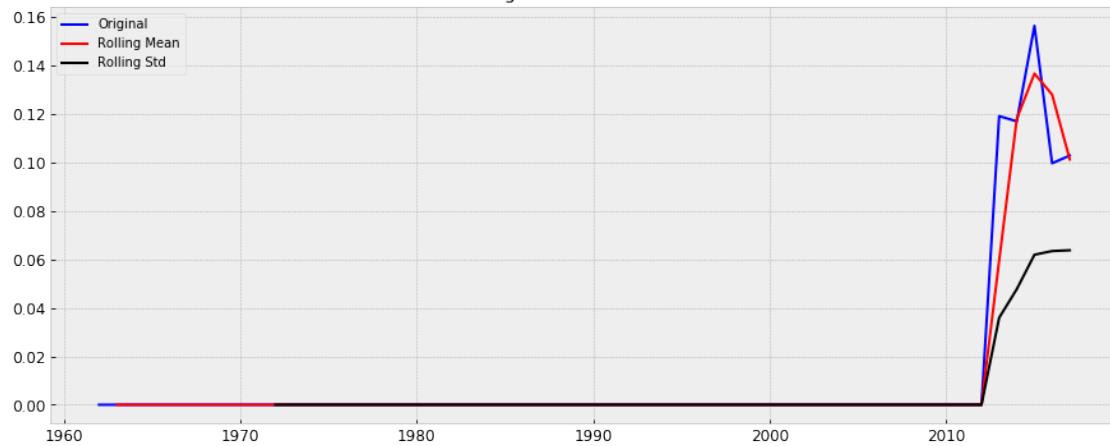
Cabo Verde: Rolling Mean & Standard Deviation



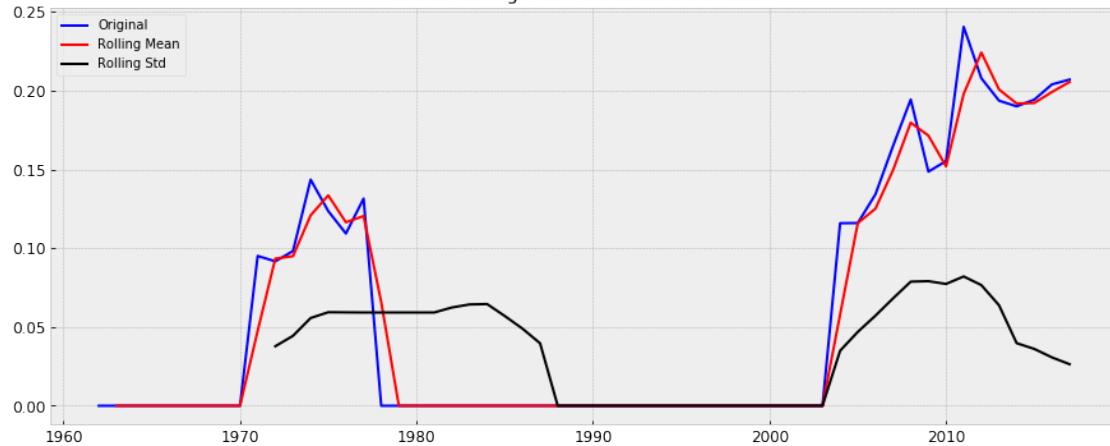




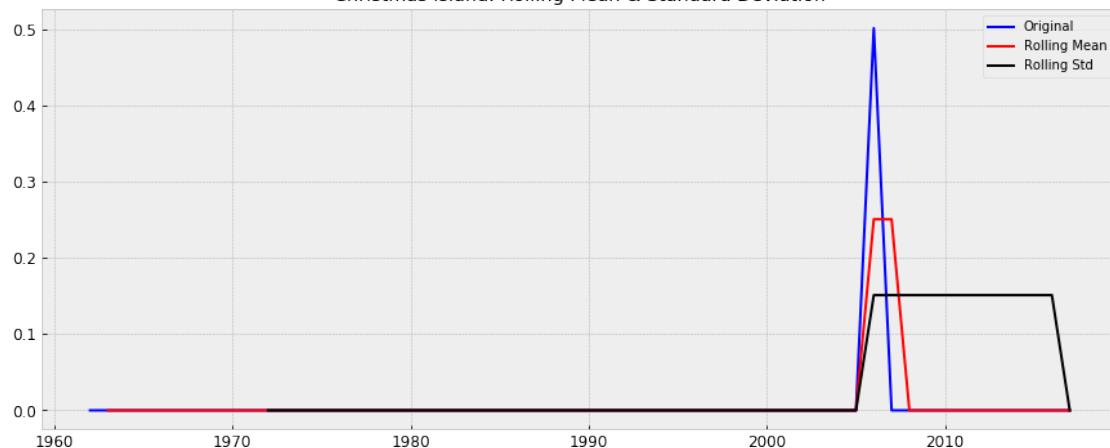
Chile: Rolling Mean & Standard Deviation



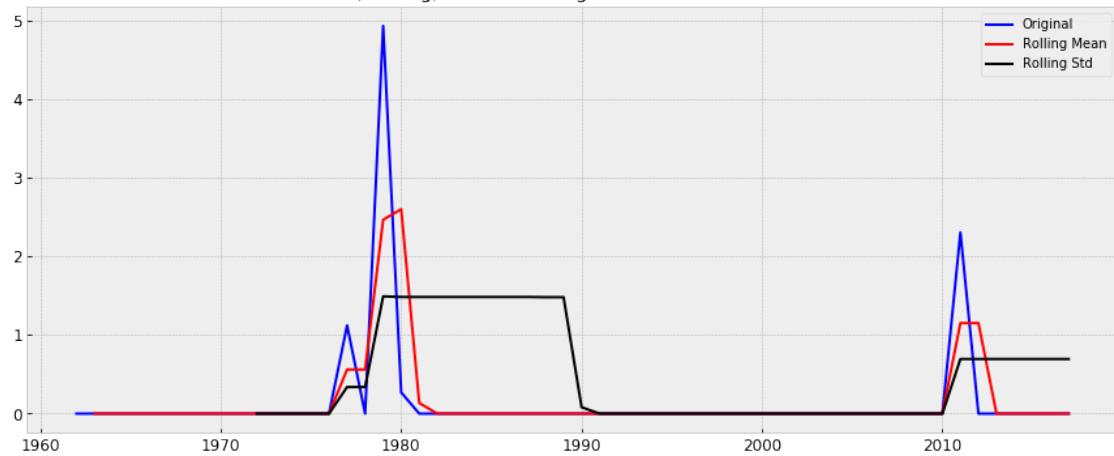
China: Rolling Mean & Standard Deviation



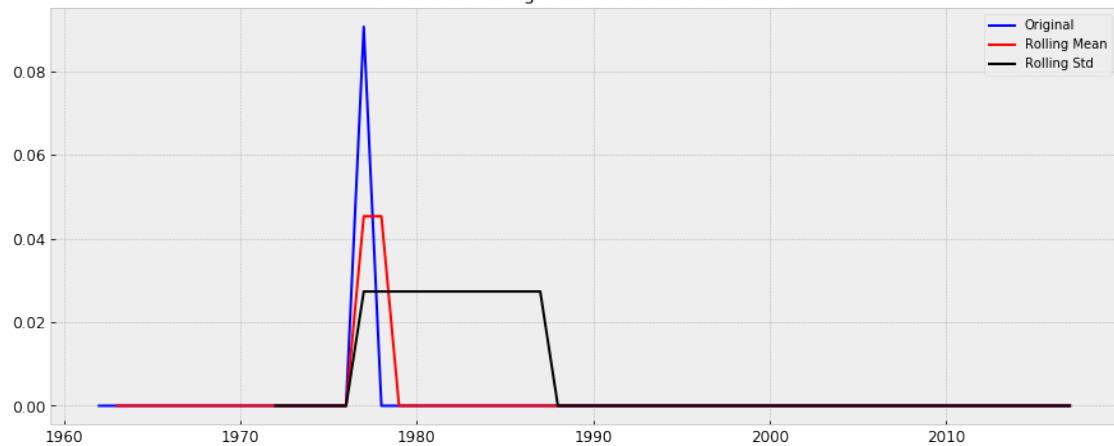
Christmas Island: Rolling Mean & Standard Deviation



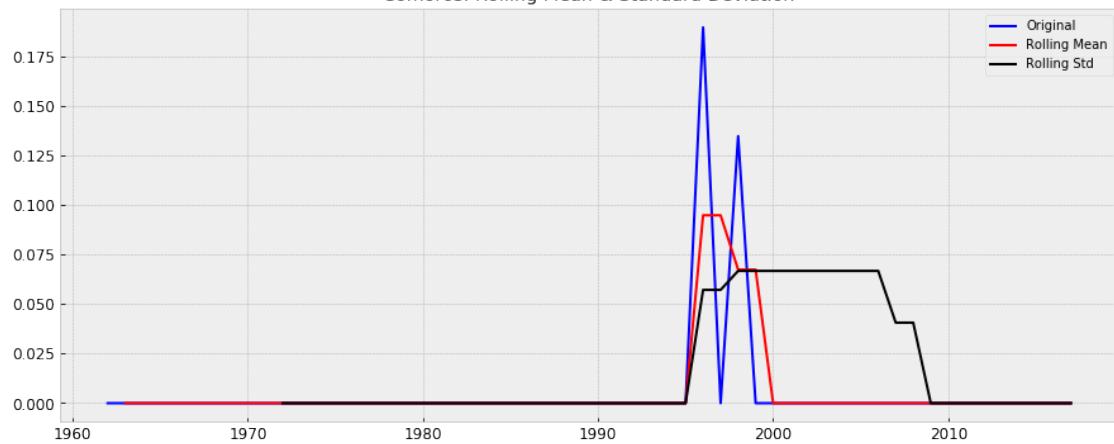
Cocos (Keeling) Islands: Rolling Mean & Standard Deviation



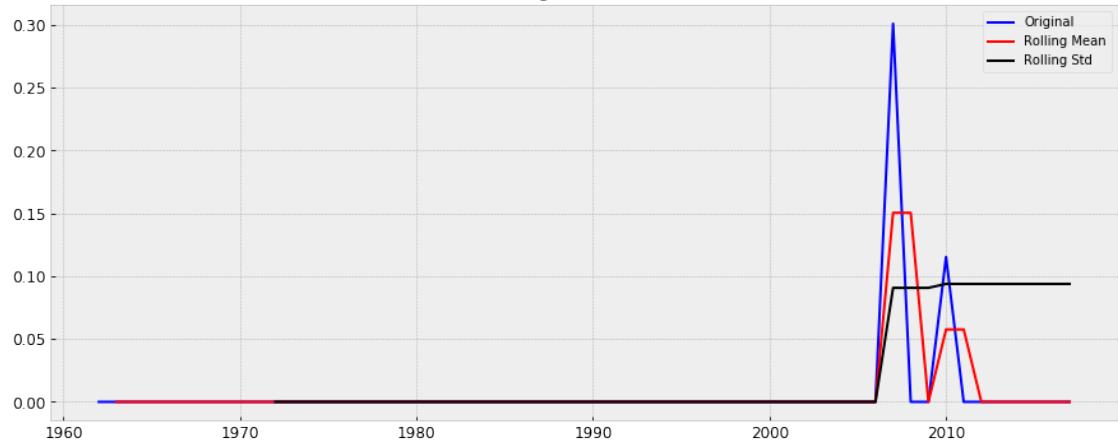
Colombia: Rolling Mean & Standard Deviation



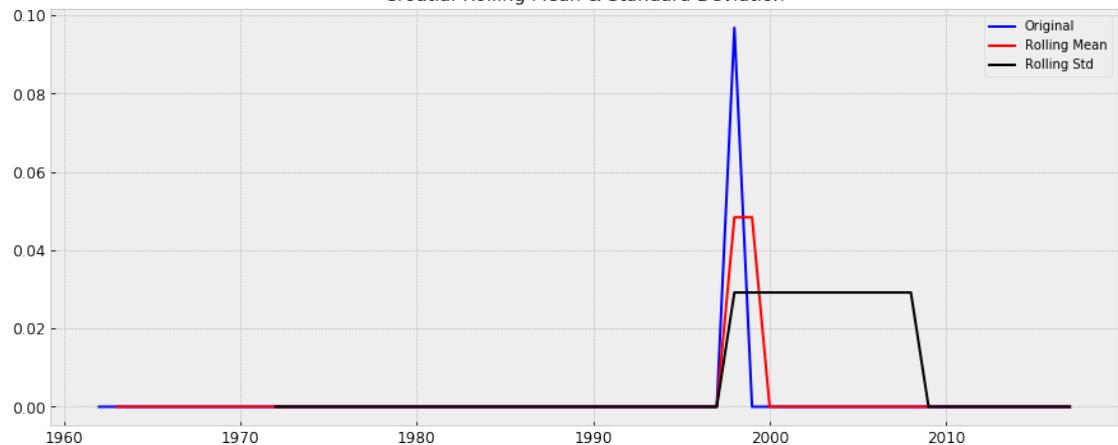
Comoros: Rolling Mean & Standard Deviation



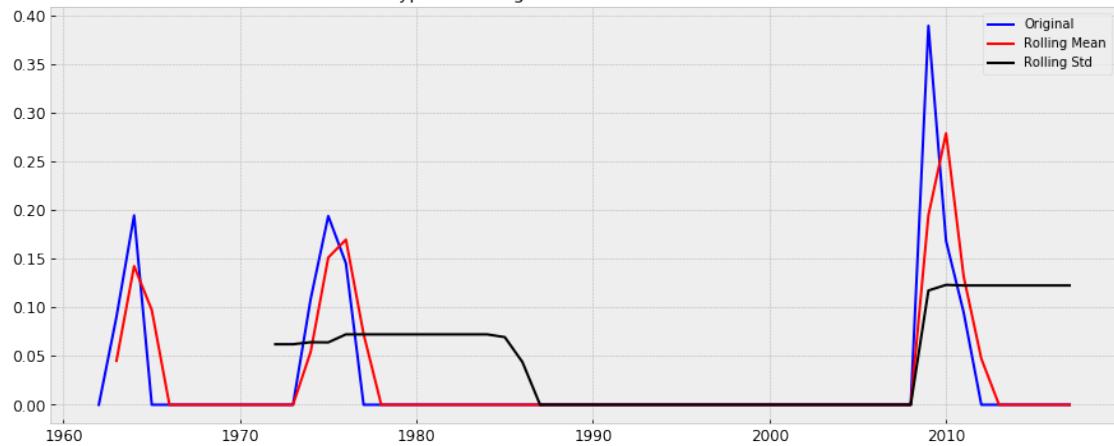
Cook Islands: Rolling Mean & Standard Deviation



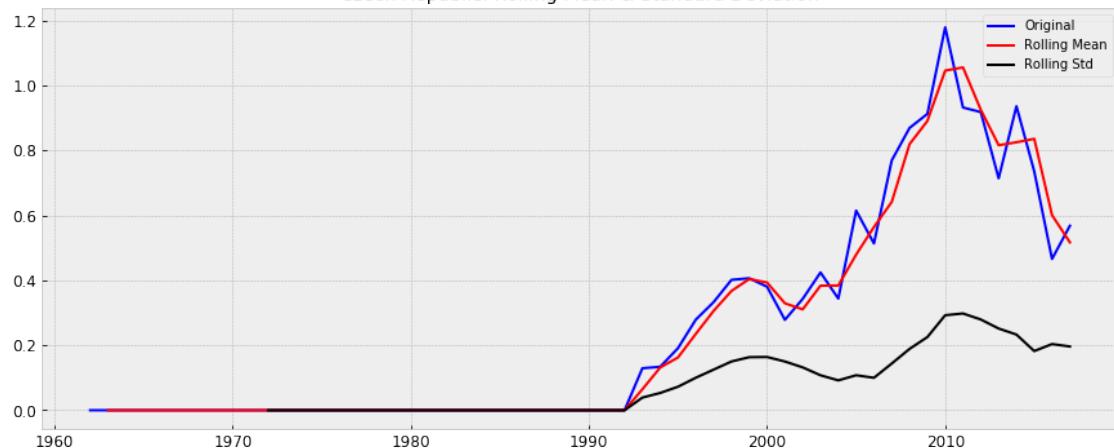
Croatia: Rolling Mean & Standard Deviation



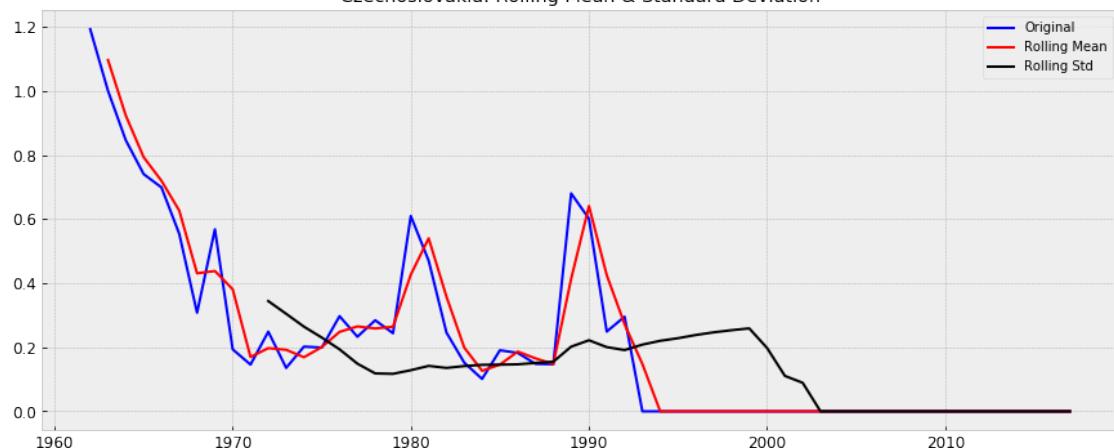
Cyprus: Rolling Mean & Standard Deviation



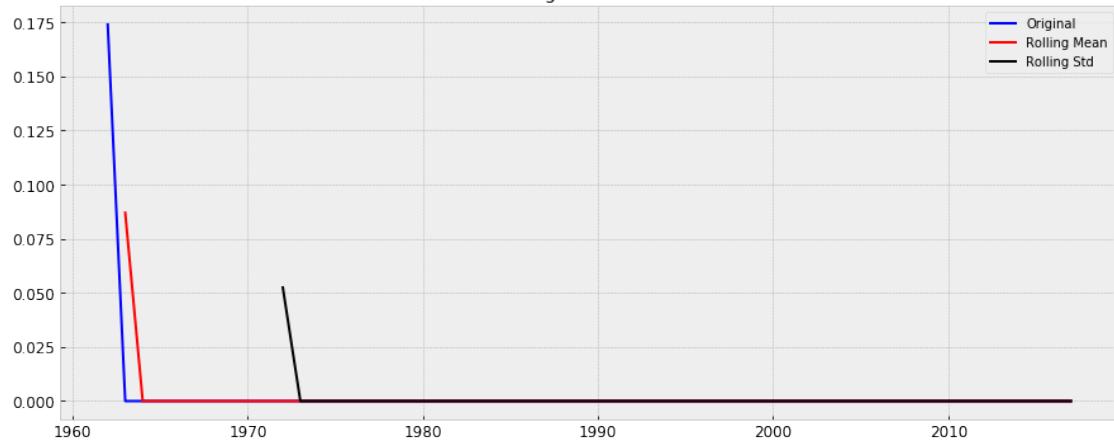
Czech Republic: Rolling Mean & Standard Deviation



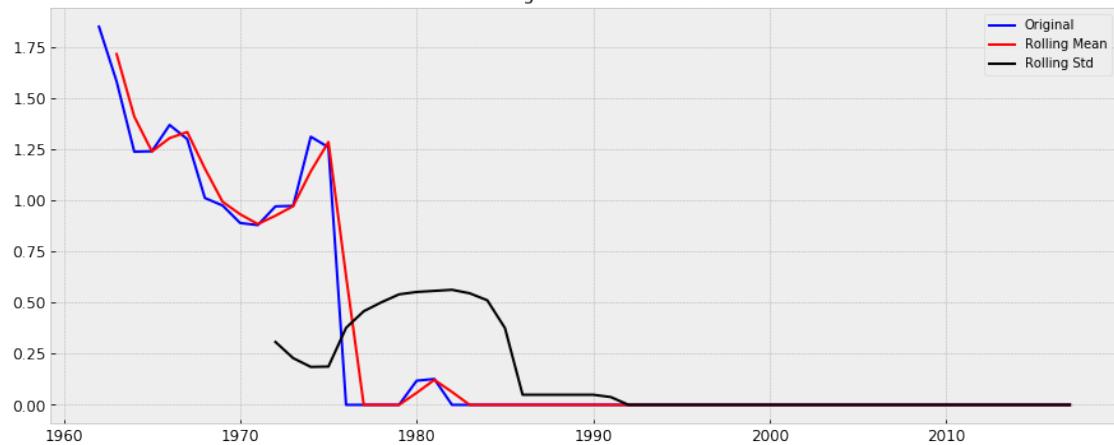
Czechoslovakia: Rolling Mean & Standard Deviation



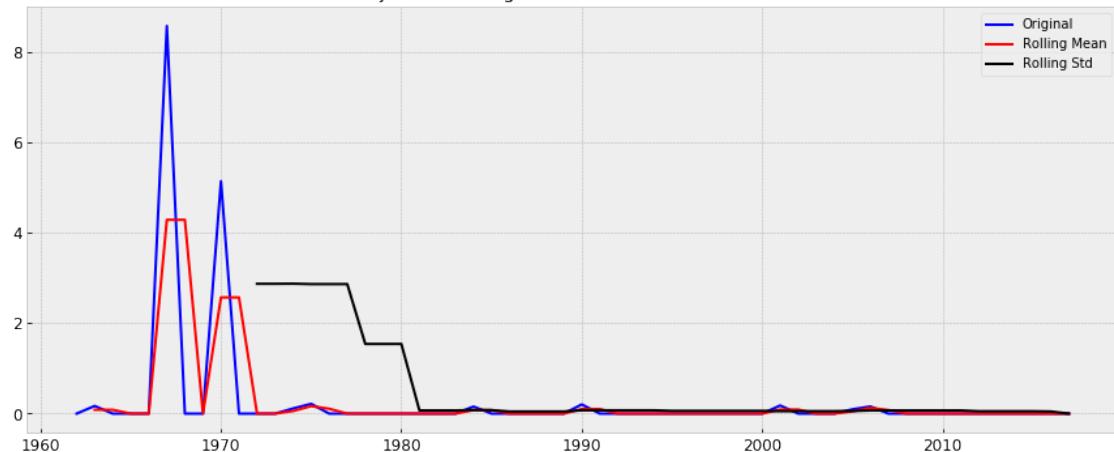
Côte d'Ivoire: Rolling Mean & Standard Deviation



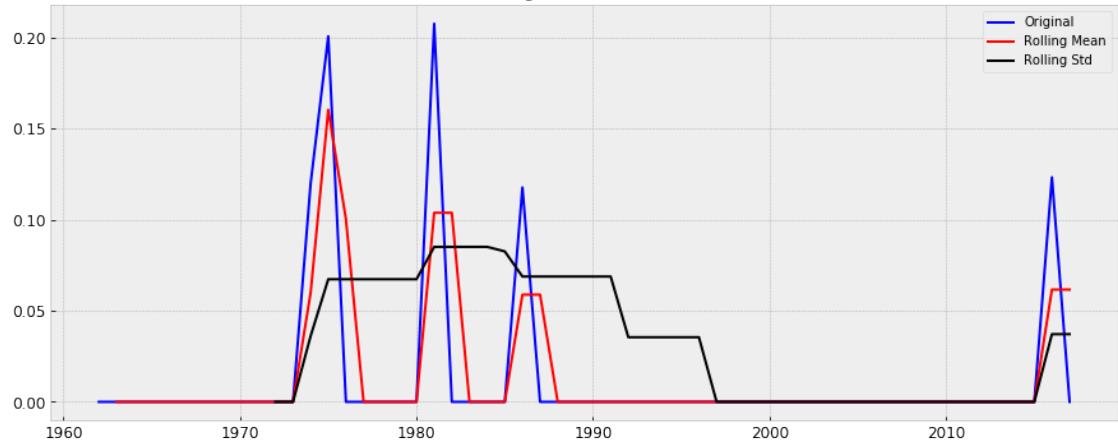
Denmark: Rolling Mean & Standard Deviation



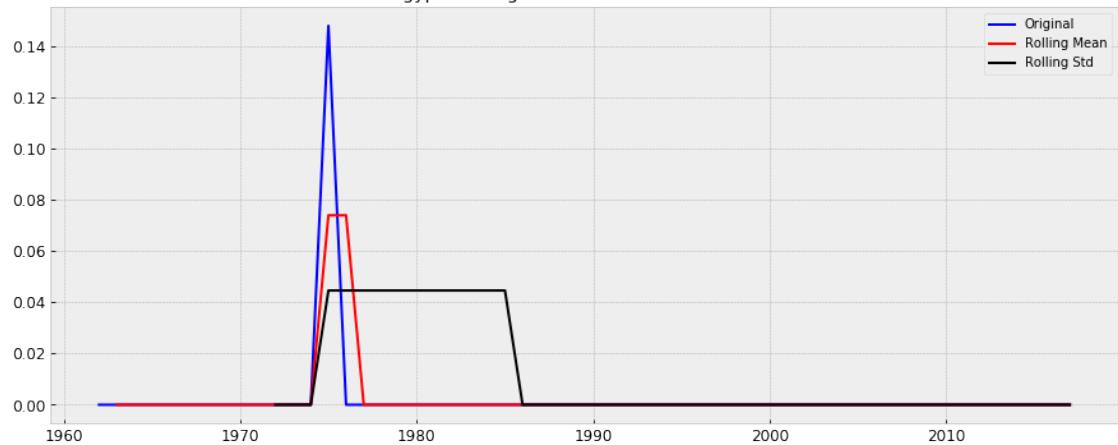
Djibouti: Rolling Mean & Standard Deviation

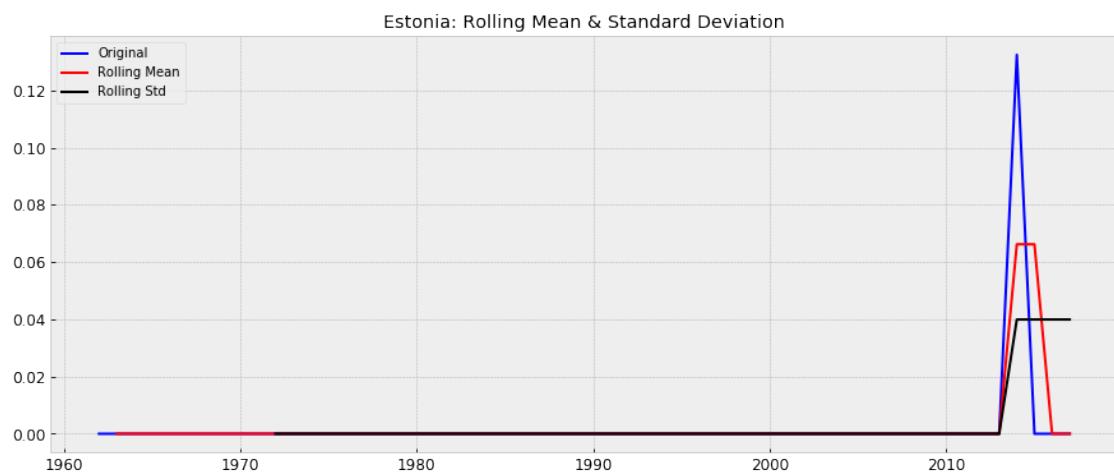
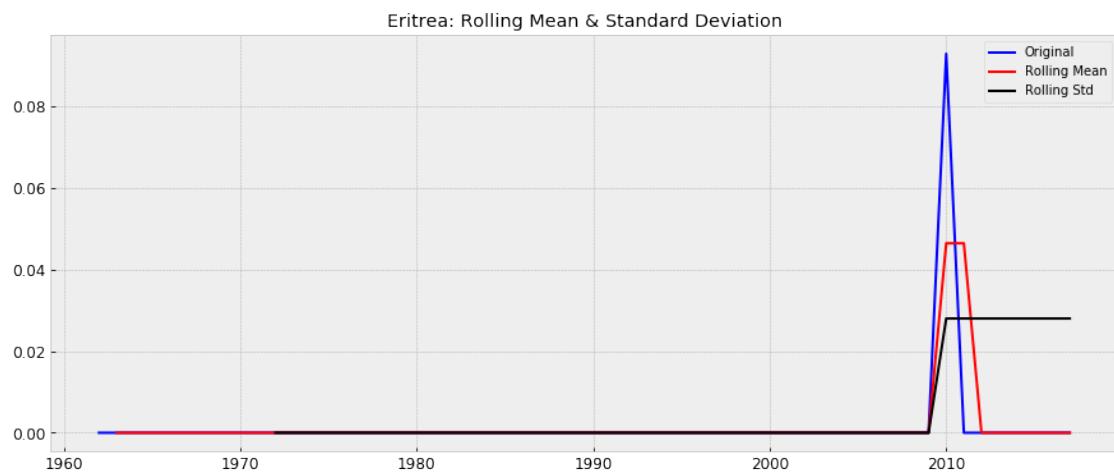
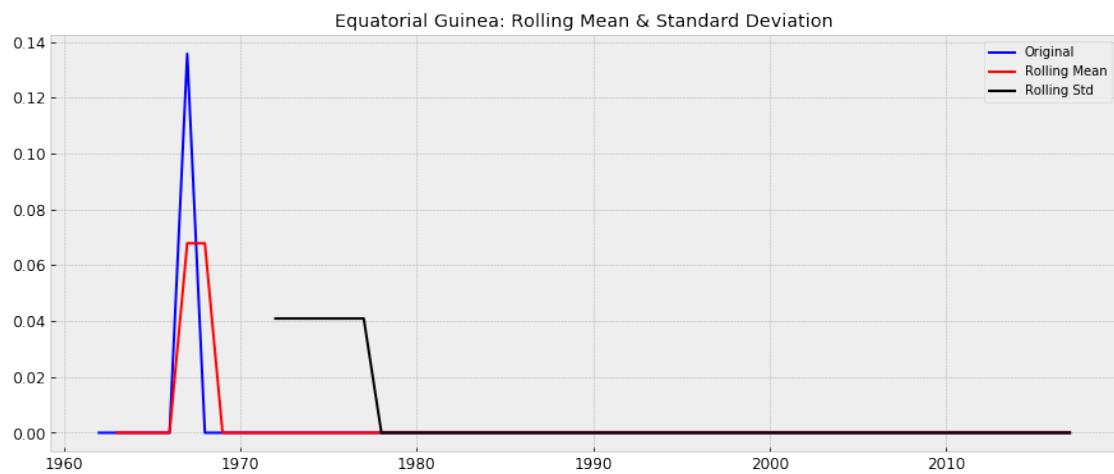


Dominica: Rolling Mean & Standard Deviation

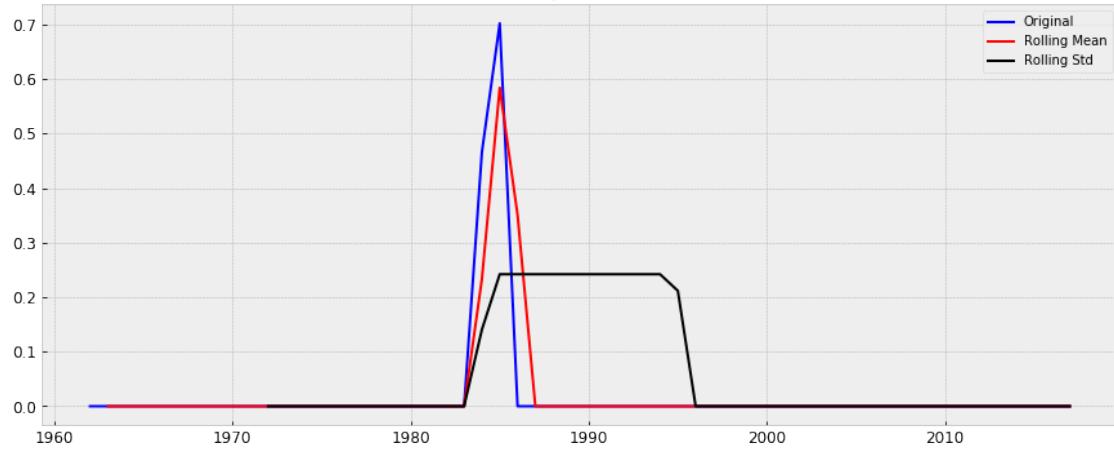


Egypt: Rolling Mean & Standard Deviation

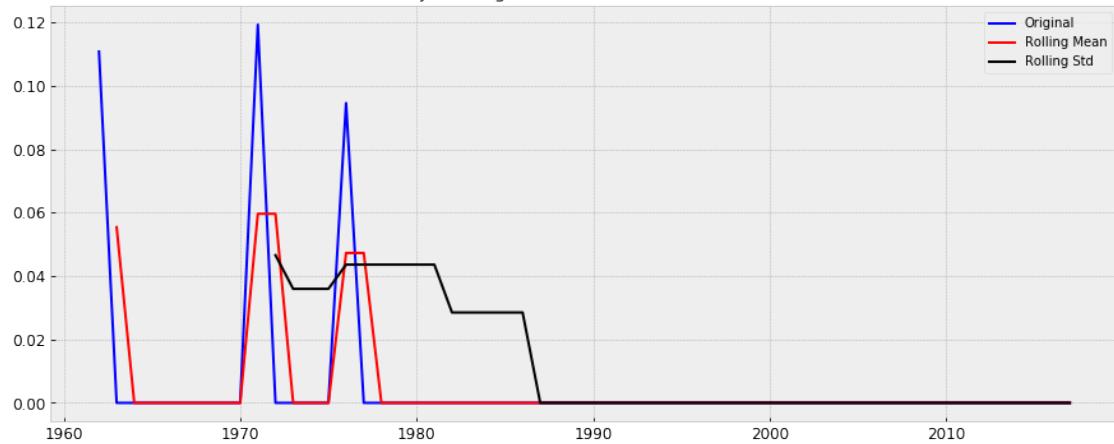




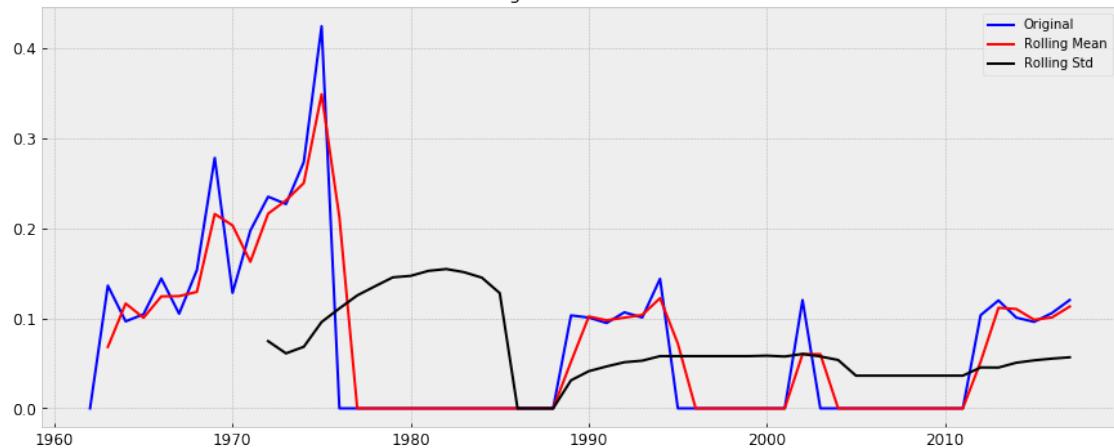
Falkland Islands: Rolling Mean & Standard Deviation



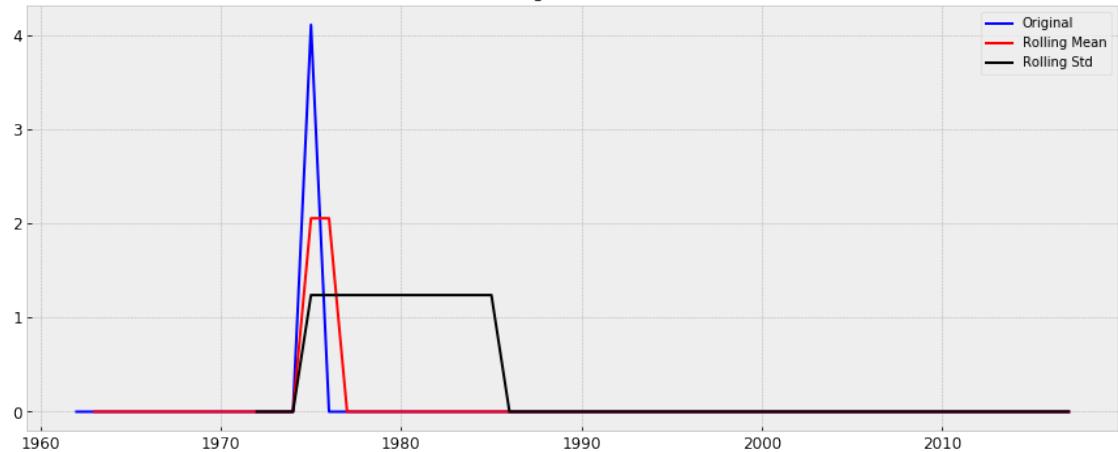
Fiji: Rolling Mean & Standard Deviation



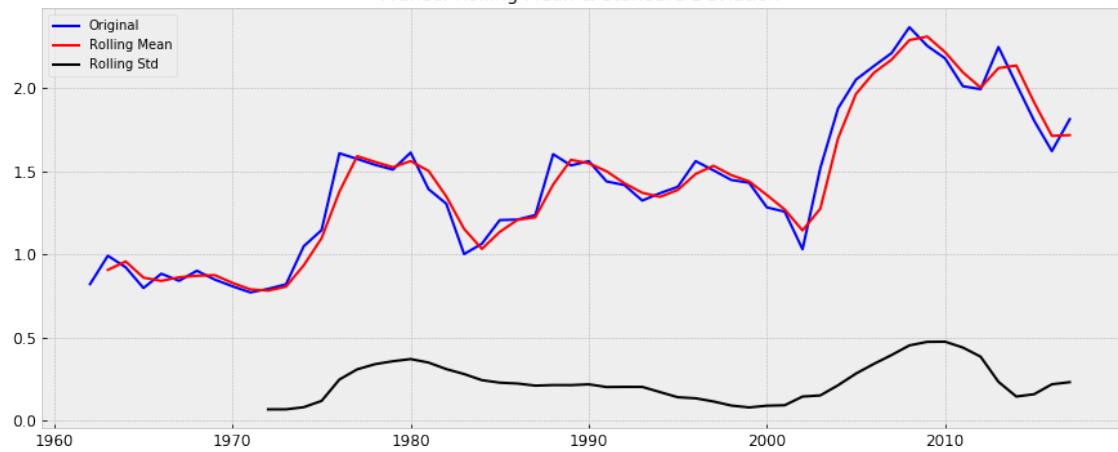
Finland: Rolling Mean & Standard Deviation



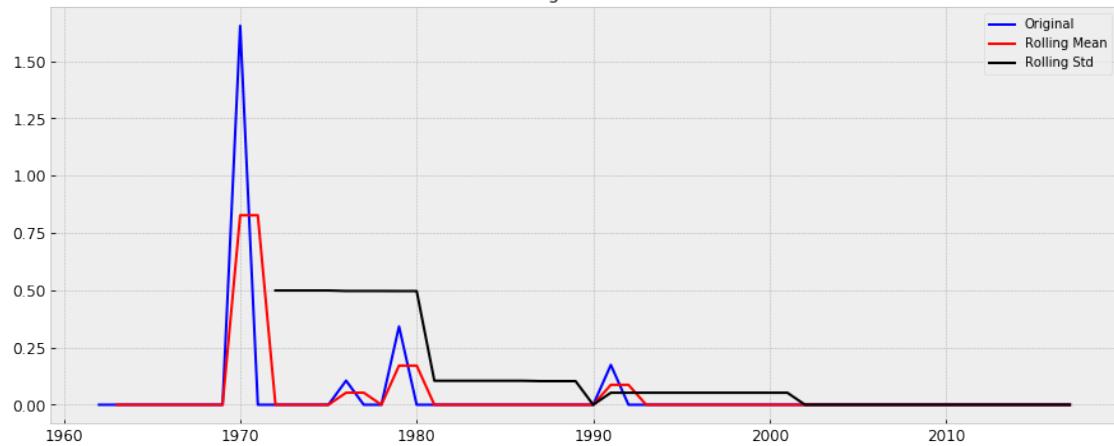
Pacific Islands: Rolling Mean & Standard Deviation



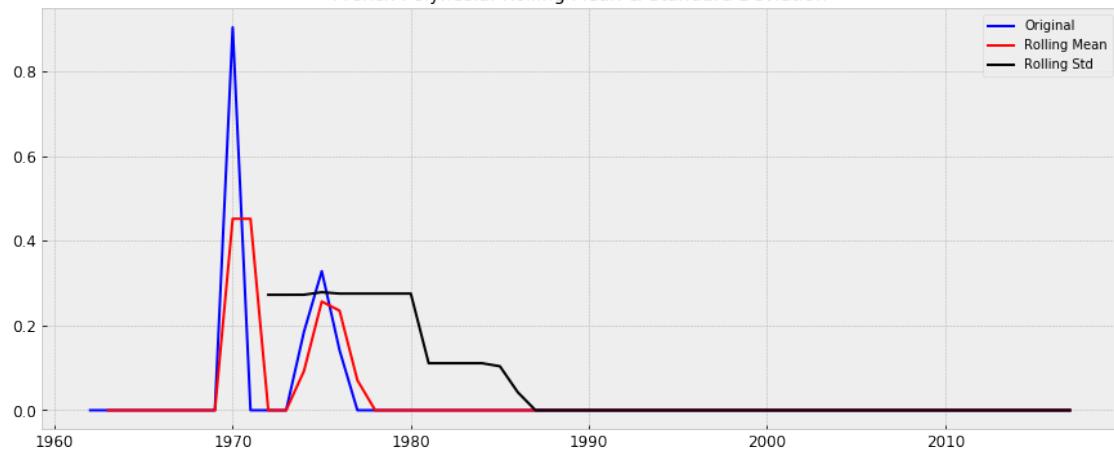
France: Rolling Mean & Standard Deviation



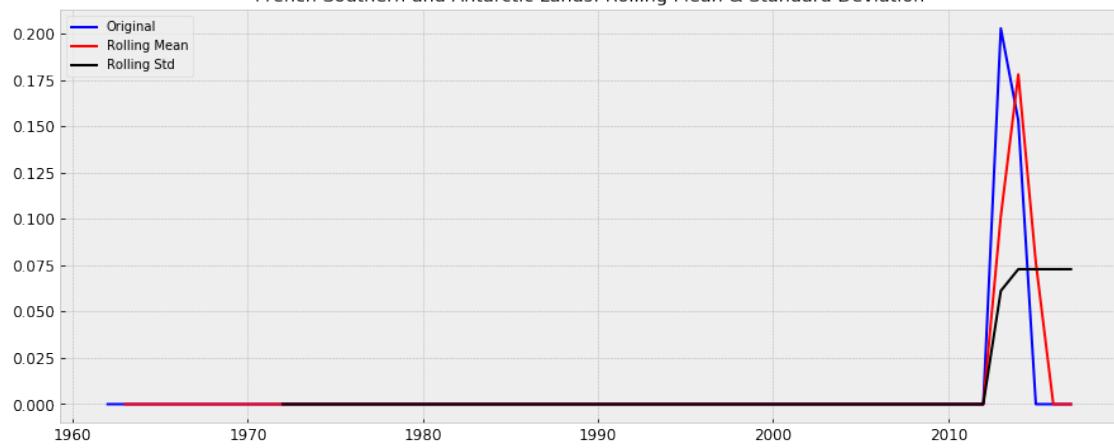
French Guiana: Rolling Mean & Standard Deviation



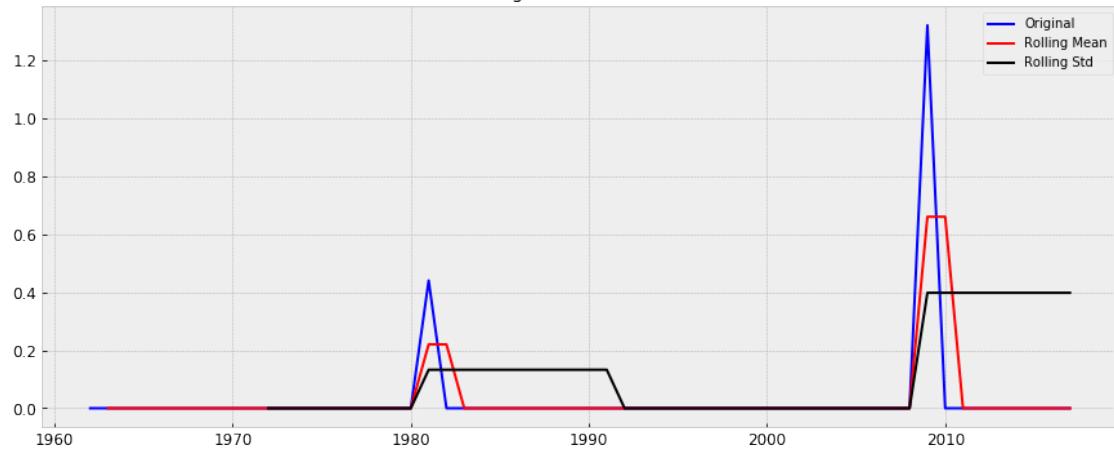
French Polynesia: Rolling Mean & Standard Deviation



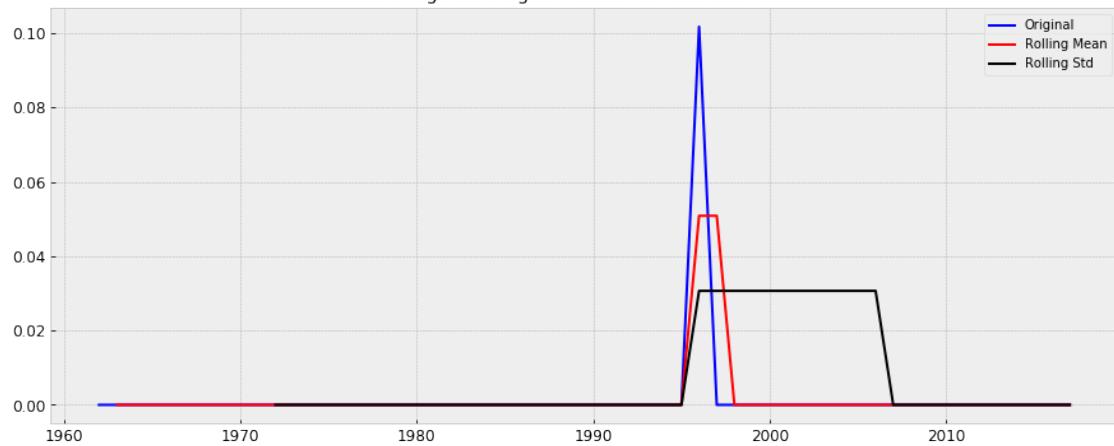
French Southern and Antarctic Lands: Rolling Mean & Standard Deviation



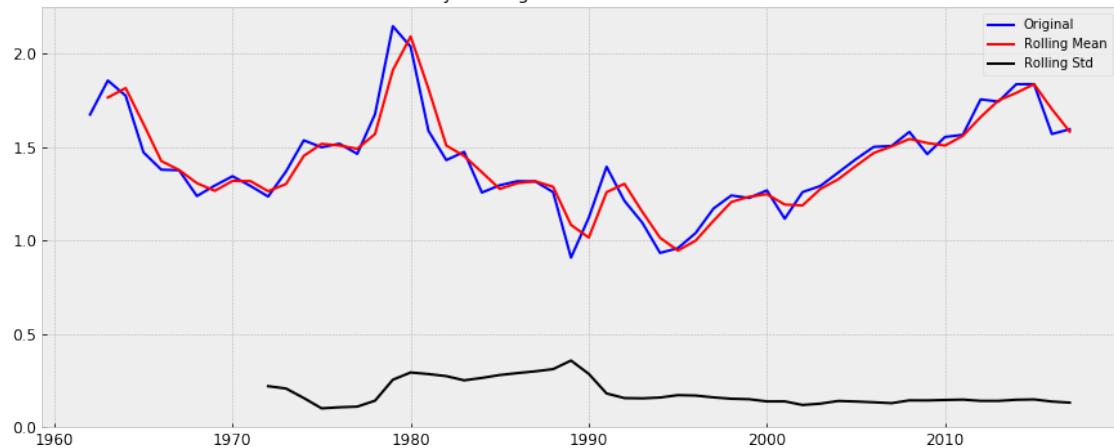
Gambia: Rolling Mean & Standard Deviation



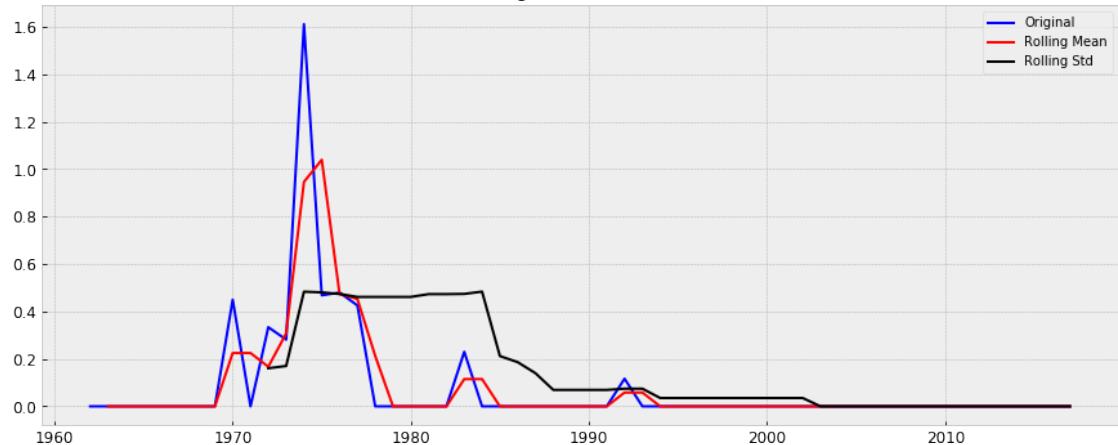
Georgia: Rolling Mean & Standard Deviation



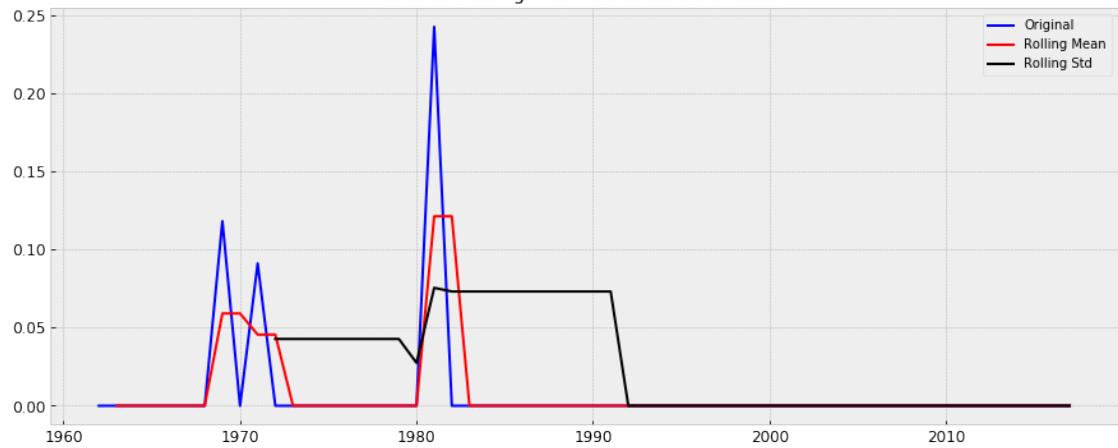
Germany: Rolling Mean & Standard Deviation



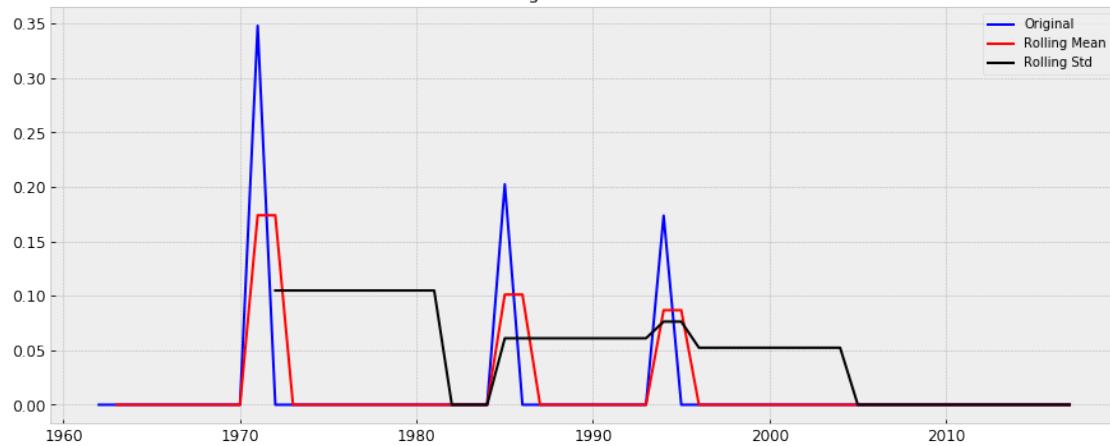
Gibraltar: Rolling Mean & Standard Deviation



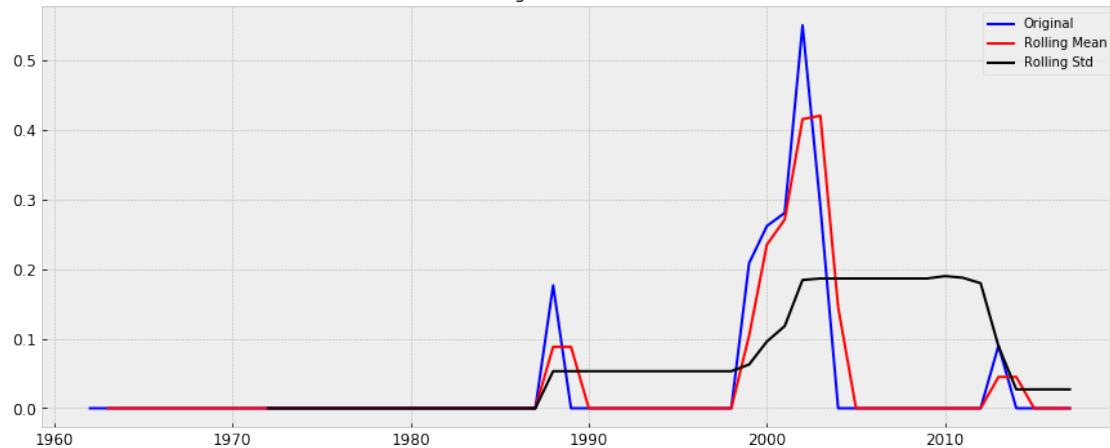
Greece: Rolling Mean & Standard Deviation



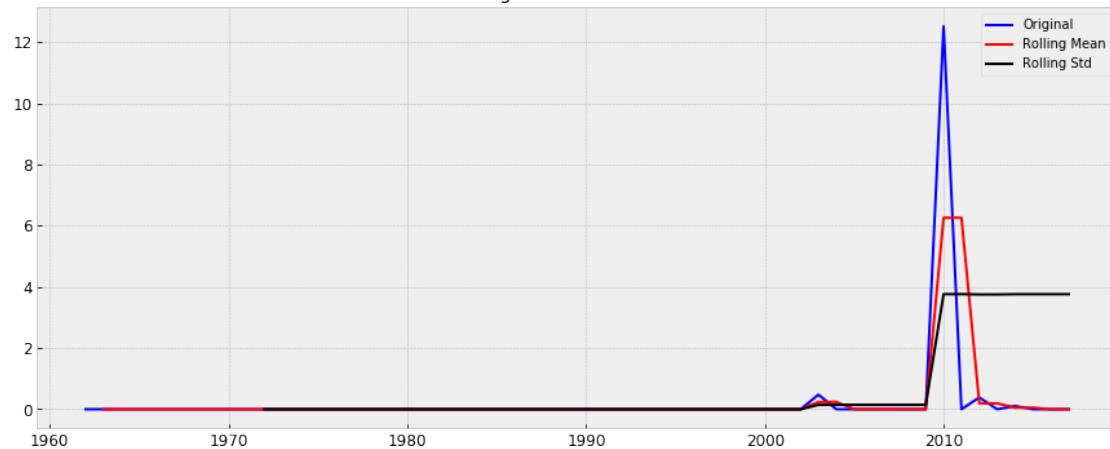
Greenland: Rolling Mean & Standard Deviation

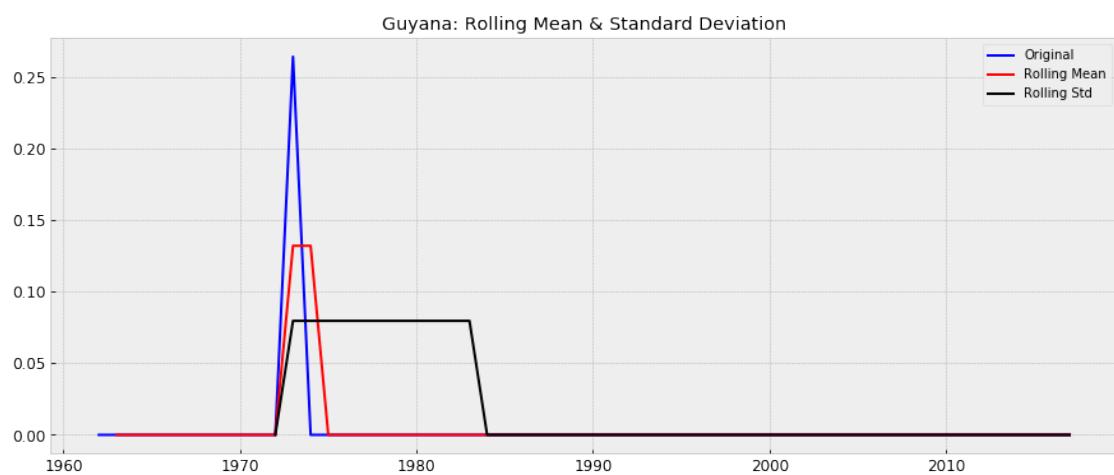
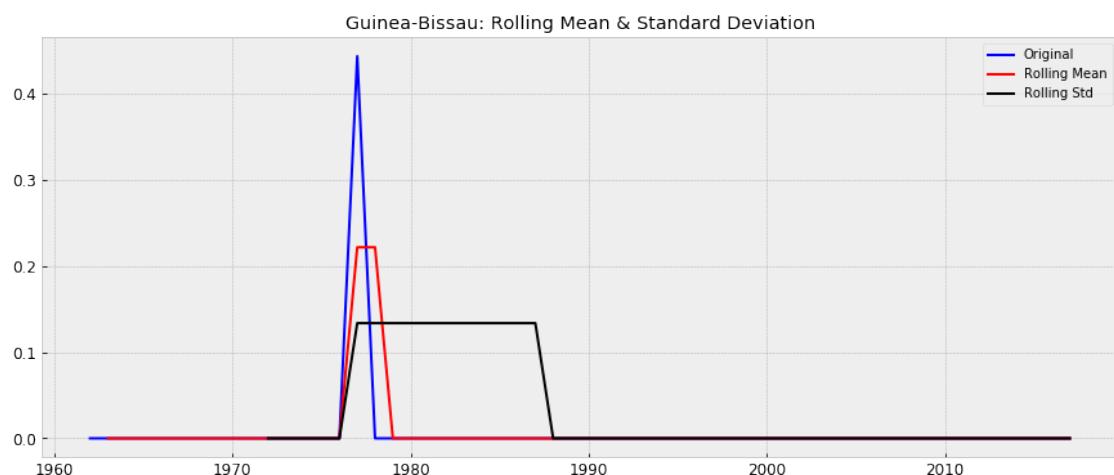
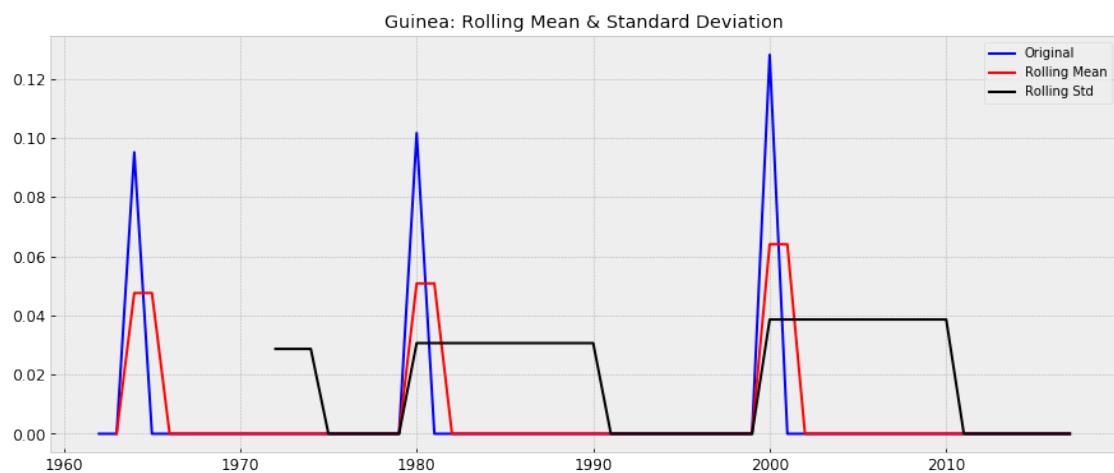


Grenada: Rolling Mean & Standard Deviation

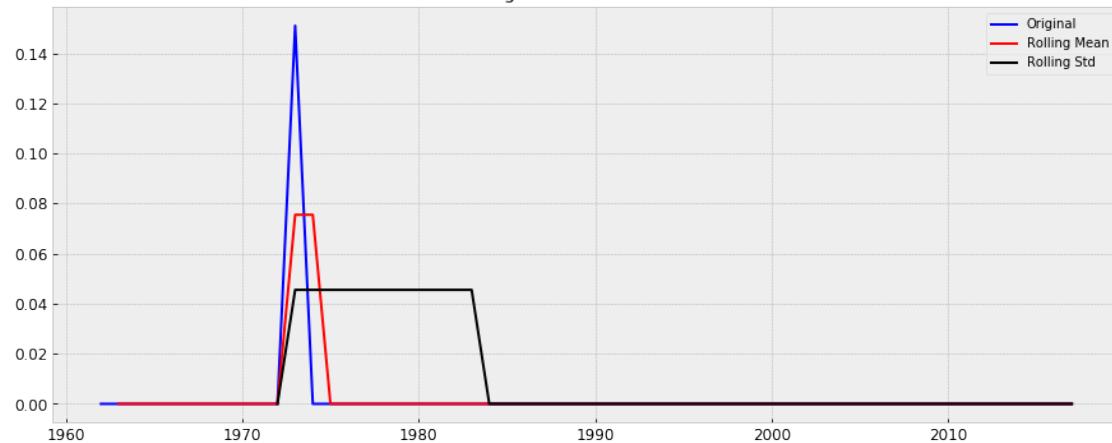


Guam: Rolling Mean & Standard Deviation

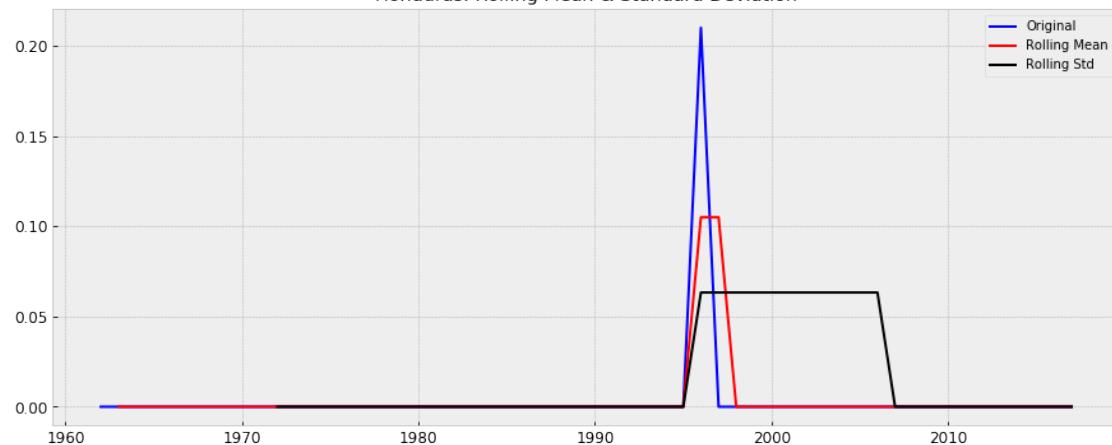




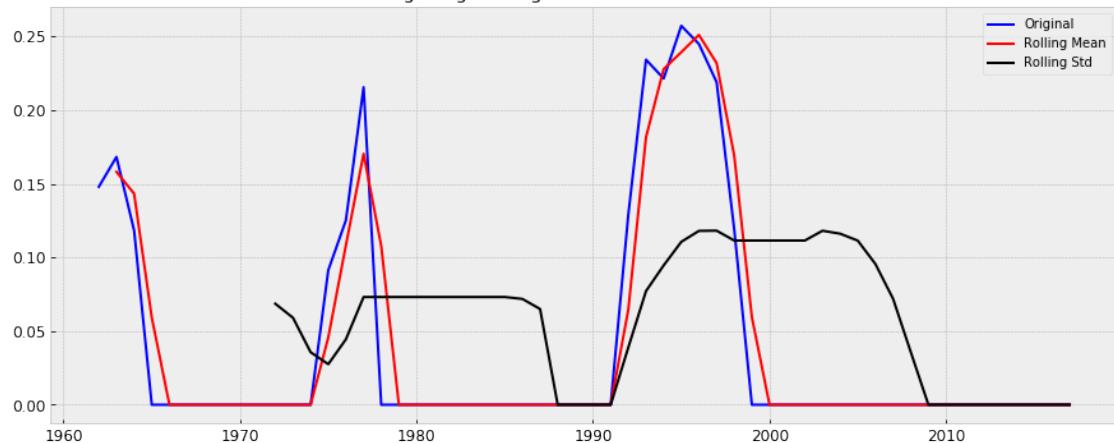
Haiti: Rolling Mean & Standard Deviation



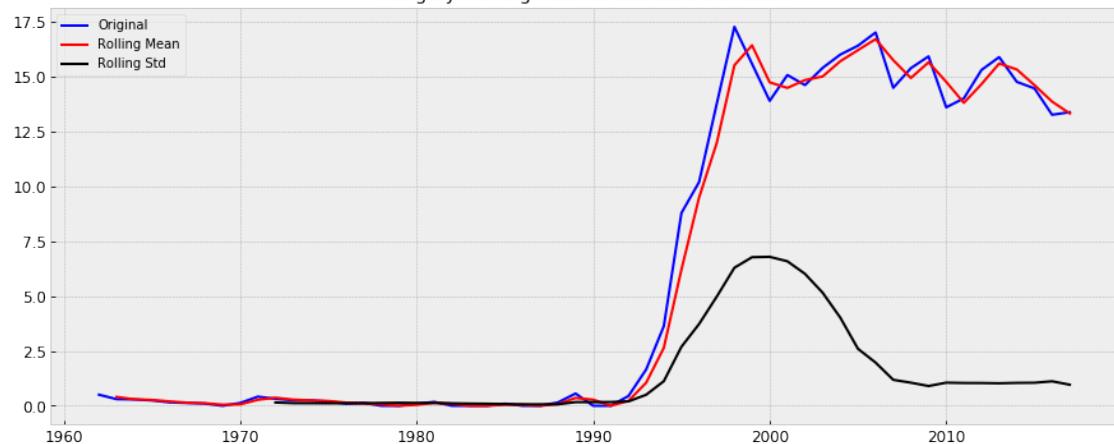
Honduras: Rolling Mean & Standard Deviation



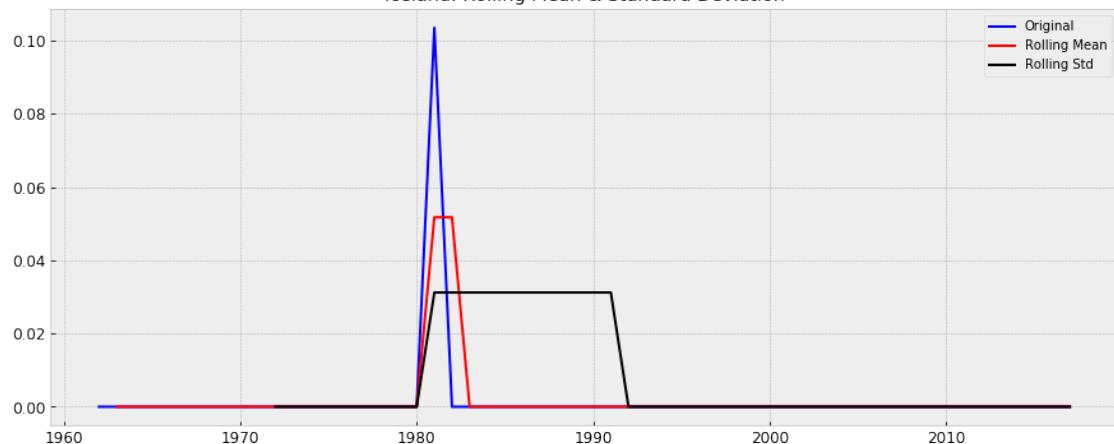
Hong Kong: Rolling Mean & Standard Deviation

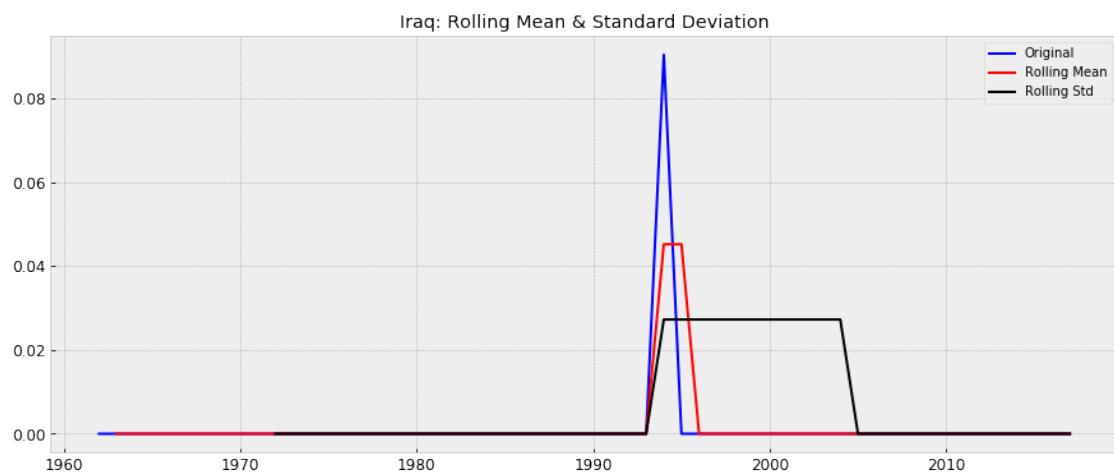
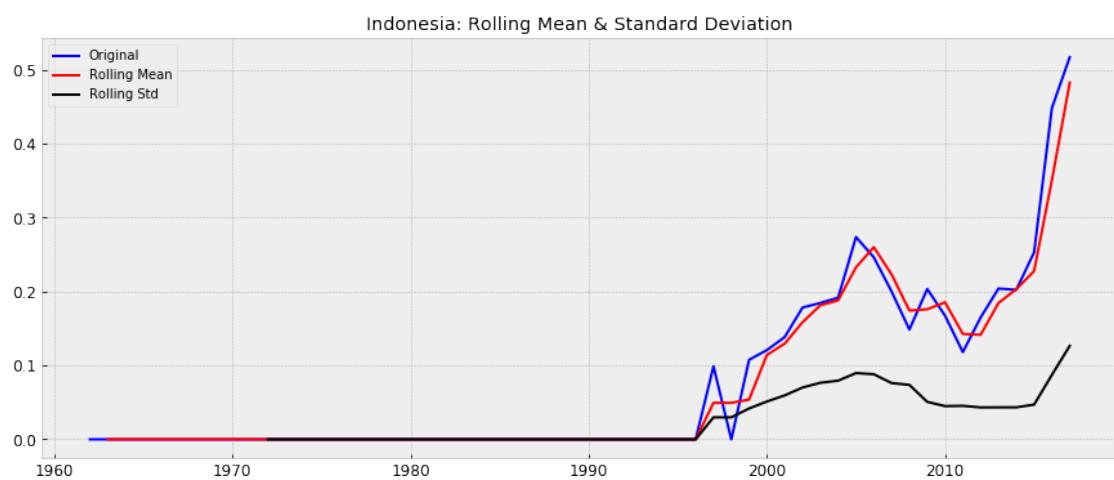
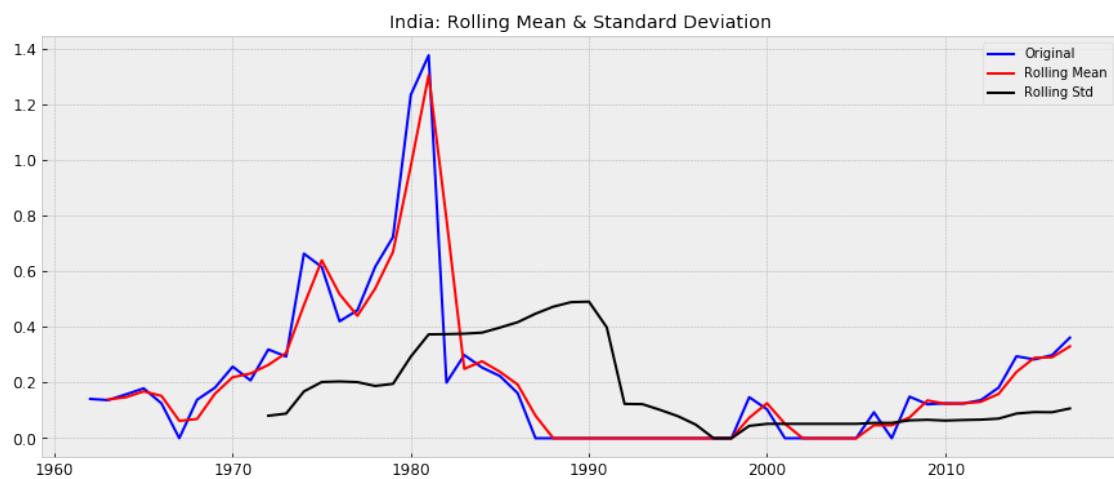


Hungary: Rolling Mean & Standard Deviation

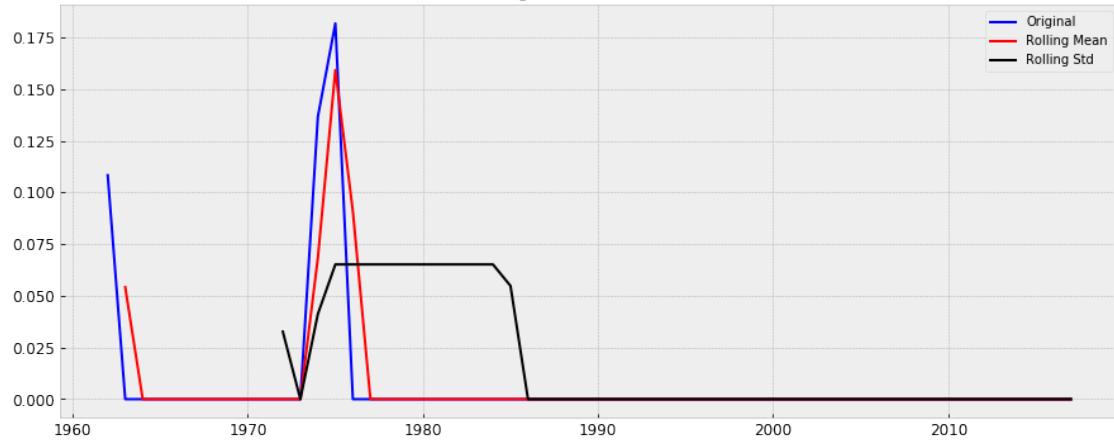


Iceland: Rolling Mean & Standard Deviation

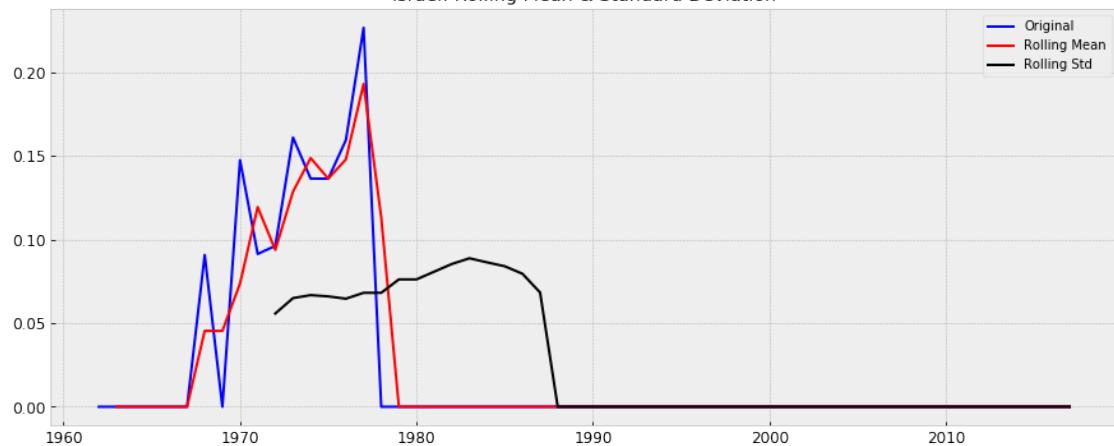




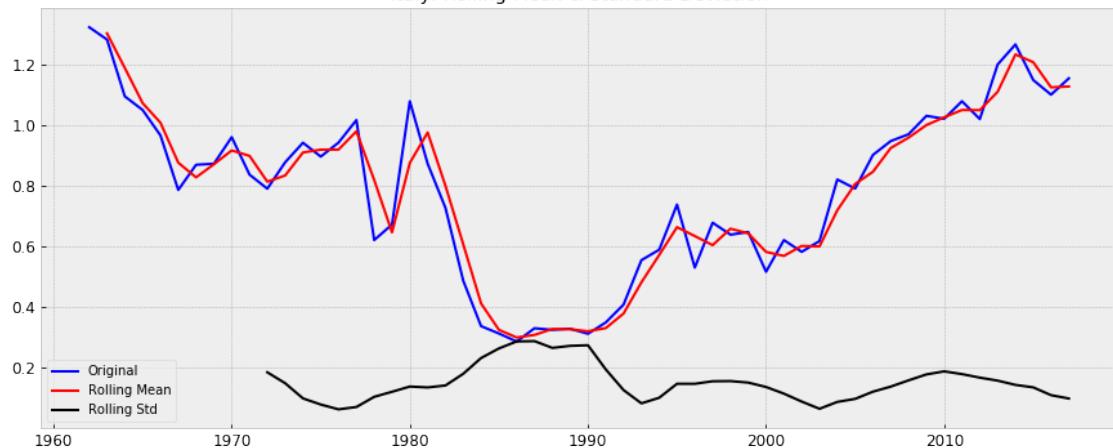
Ireland: Rolling Mean & Standard Deviation



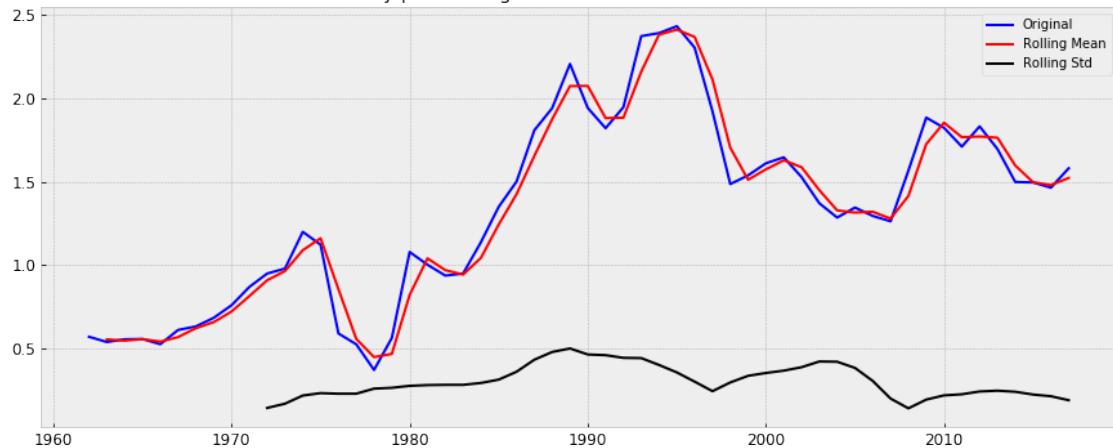
Israel: Rolling Mean & Standard Deviation



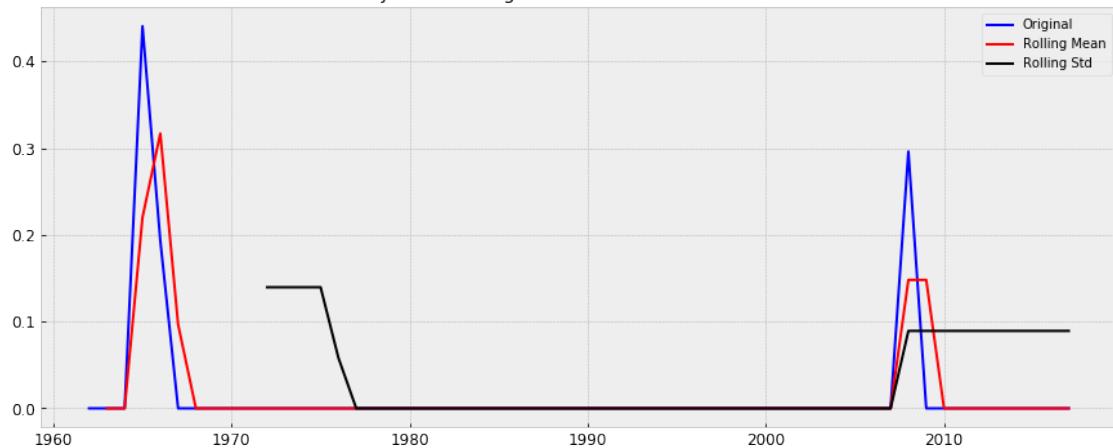
Italy: Rolling Mean & Standard Deviation

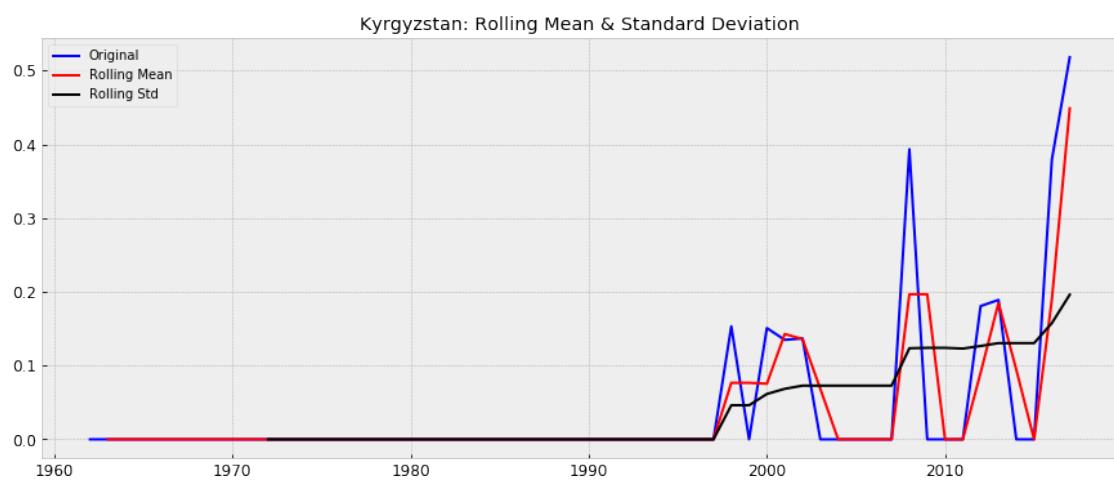
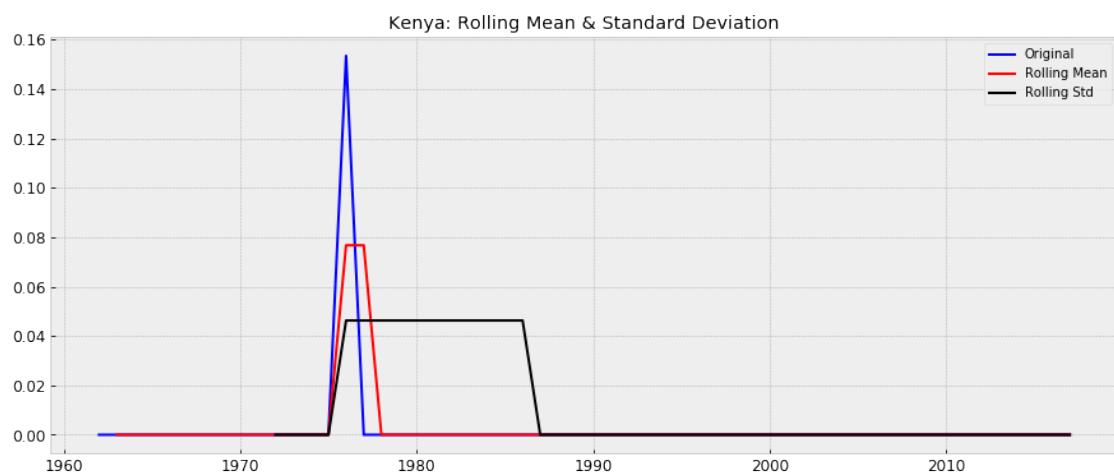
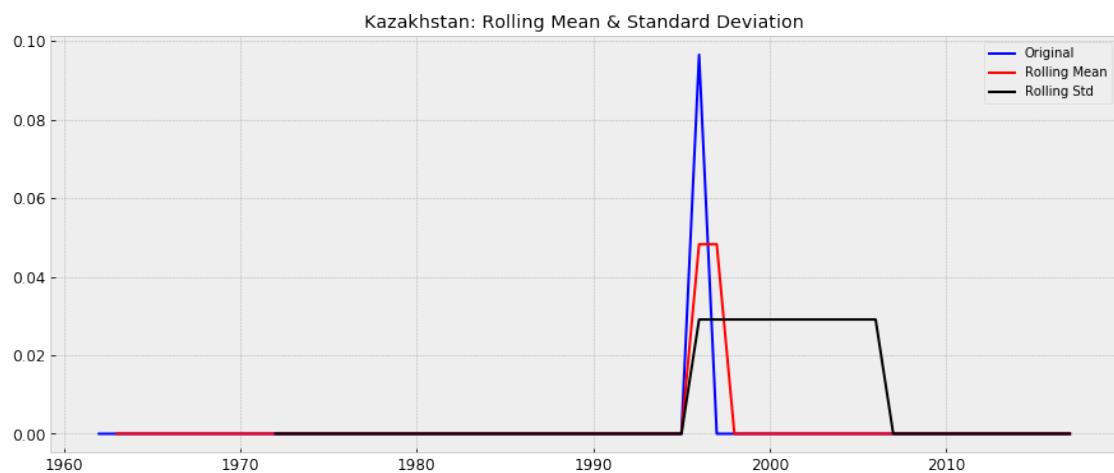


Japan: Rolling Mean & Standard Deviation

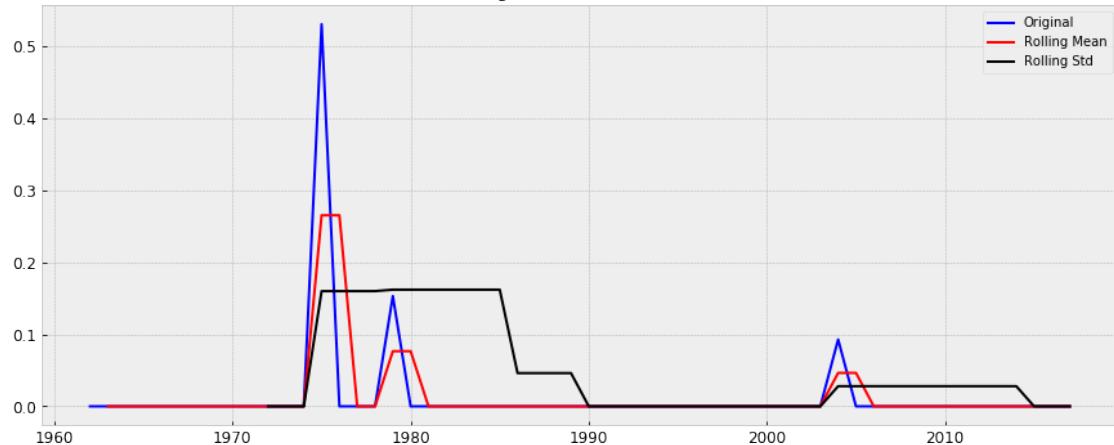


Jordan: Rolling Mean & Standard Deviation

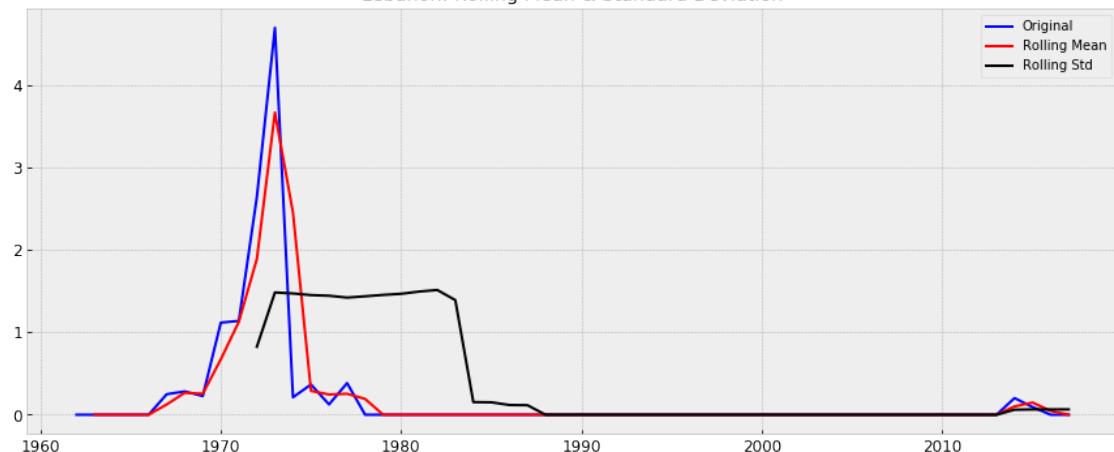




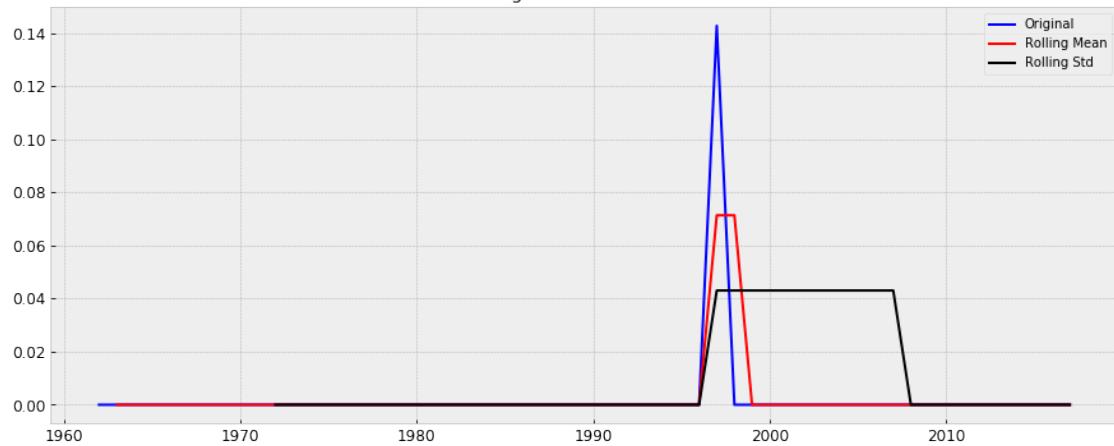
Laos: Rolling Mean & Standard Deviation



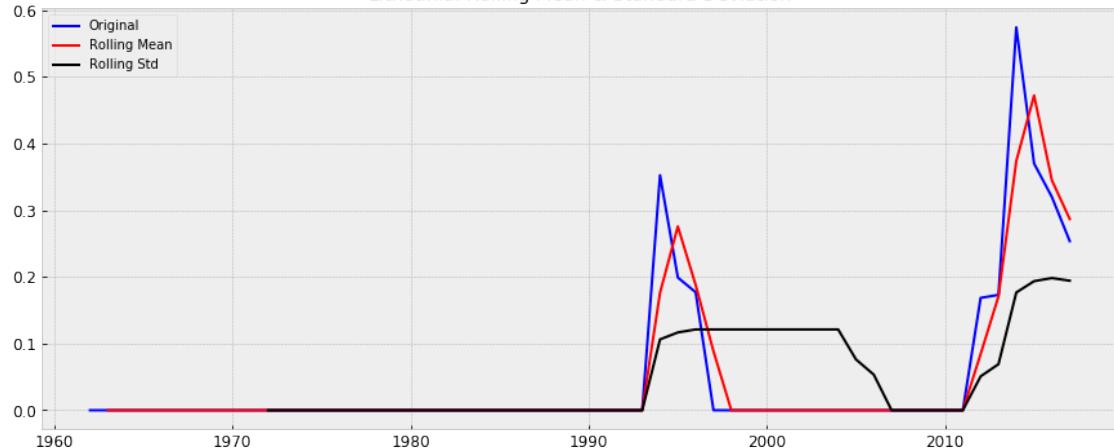
Lebanon: Rolling Mean & Standard Deviation



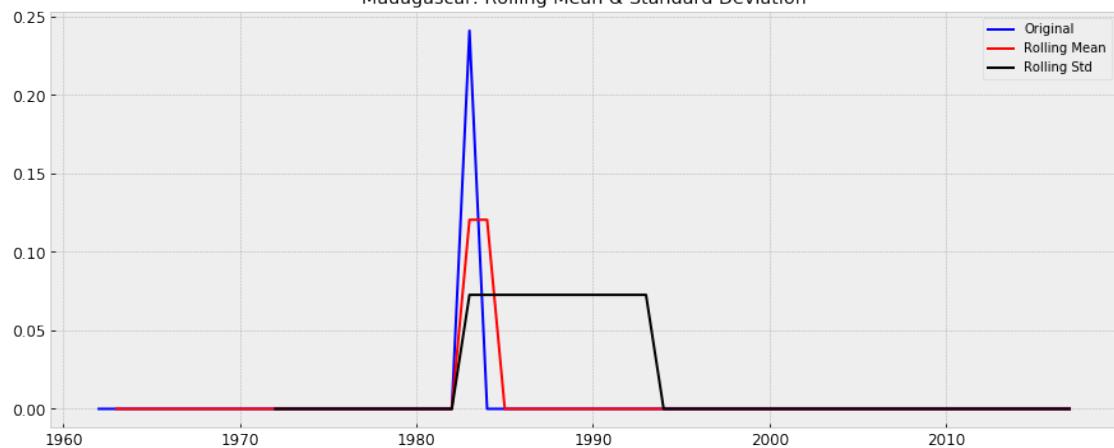
Liberia: Rolling Mean & Standard Deviation



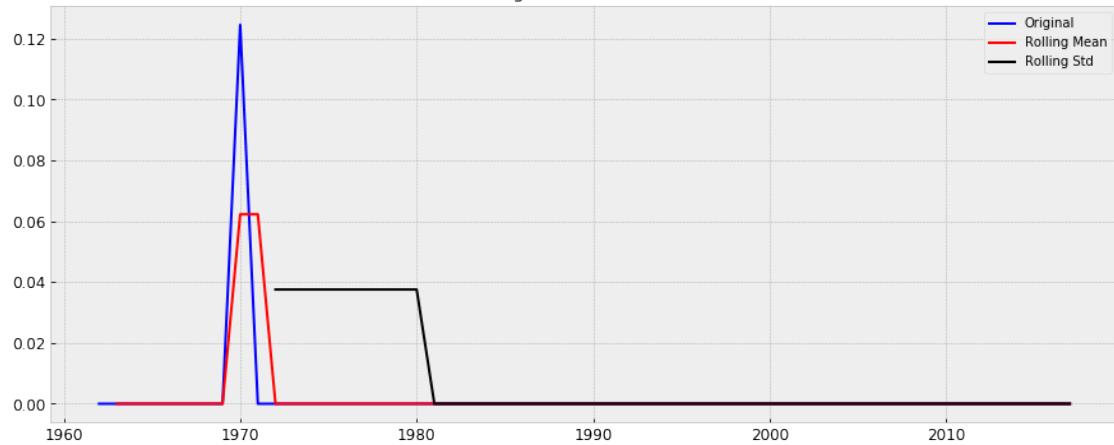
Lithuania: Rolling Mean & Standard Deviation



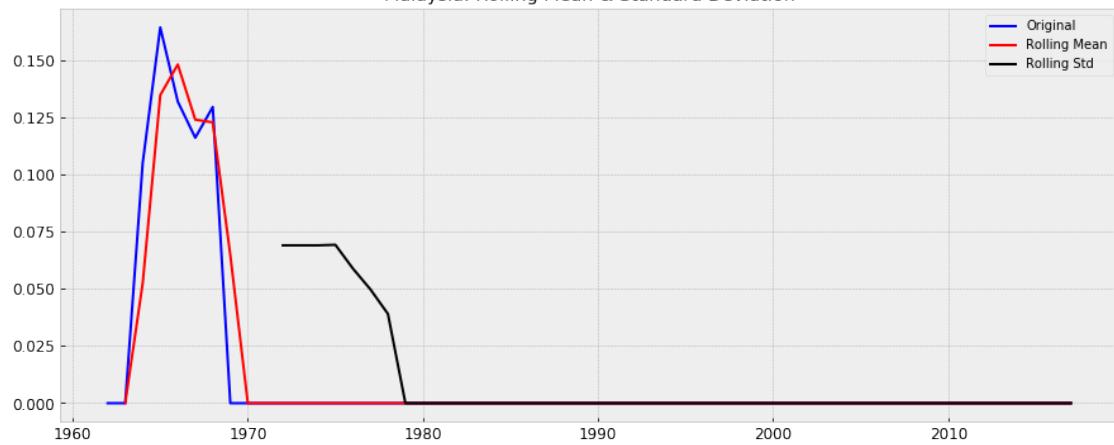
Madagascar: Rolling Mean & Standard Deviation



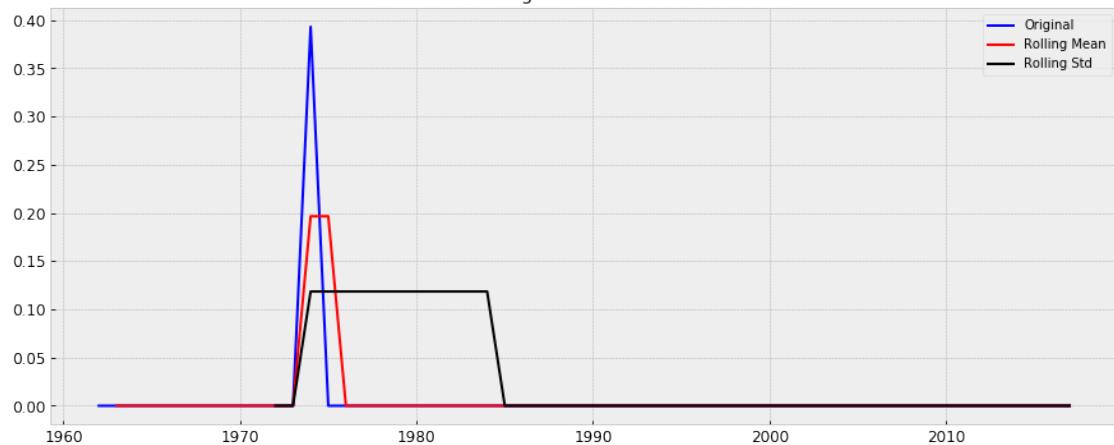
Malawi: Rolling Mean & Standard Deviation



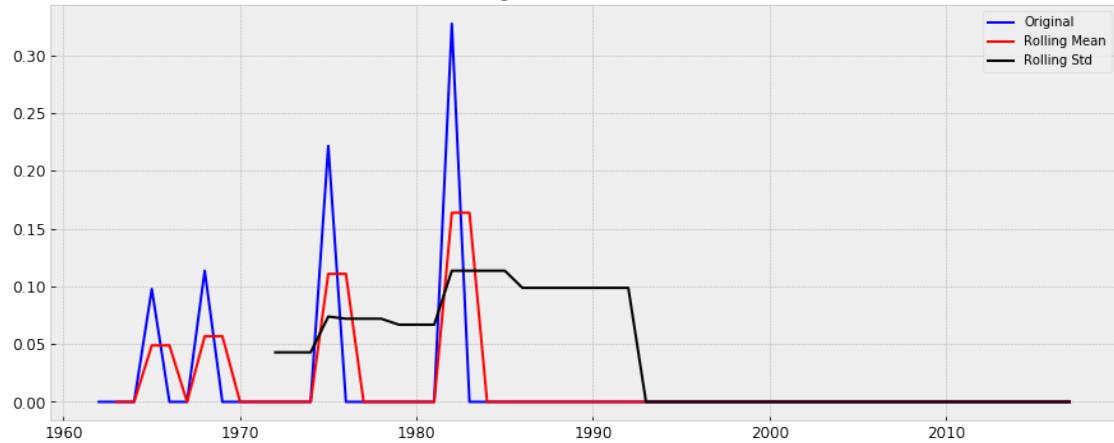
Malaysia: Rolling Mean & Standard Deviation



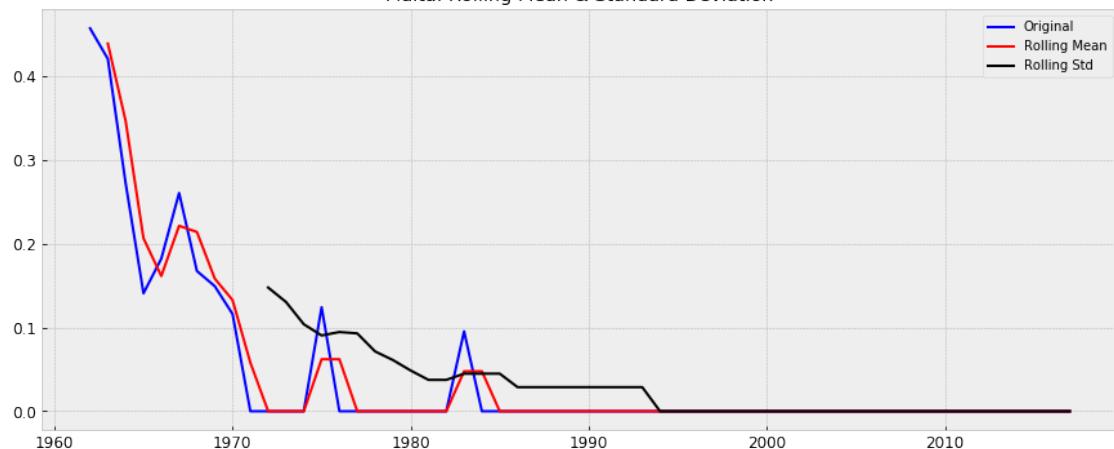
Maldives: Rolling Mean & Standard Deviation



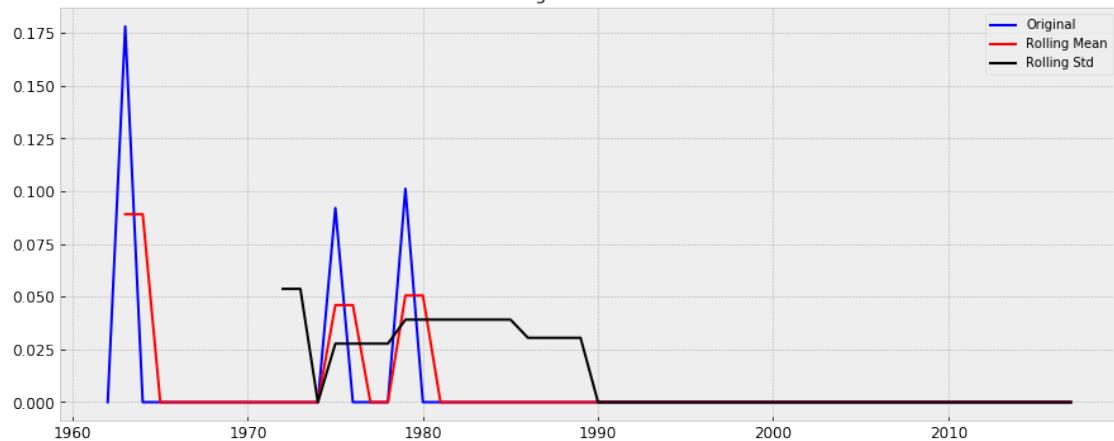
Mali: Rolling Mean & Standard Deviation



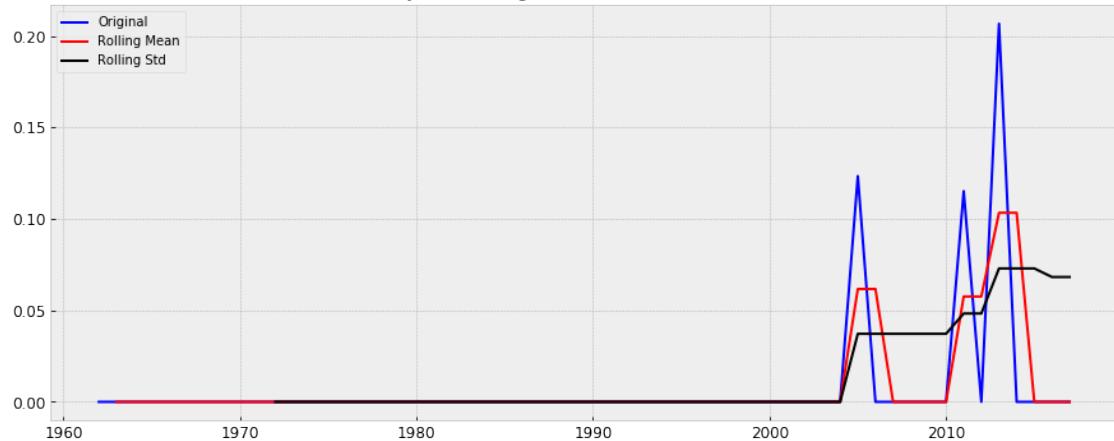
Malta: Rolling Mean & Standard Deviation



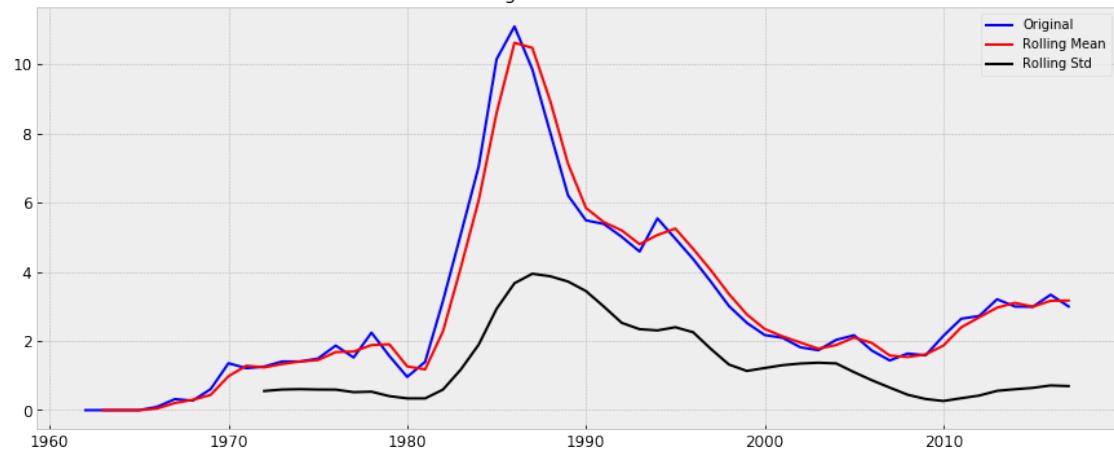
Mauritania: Rolling Mean & Standard Deviation

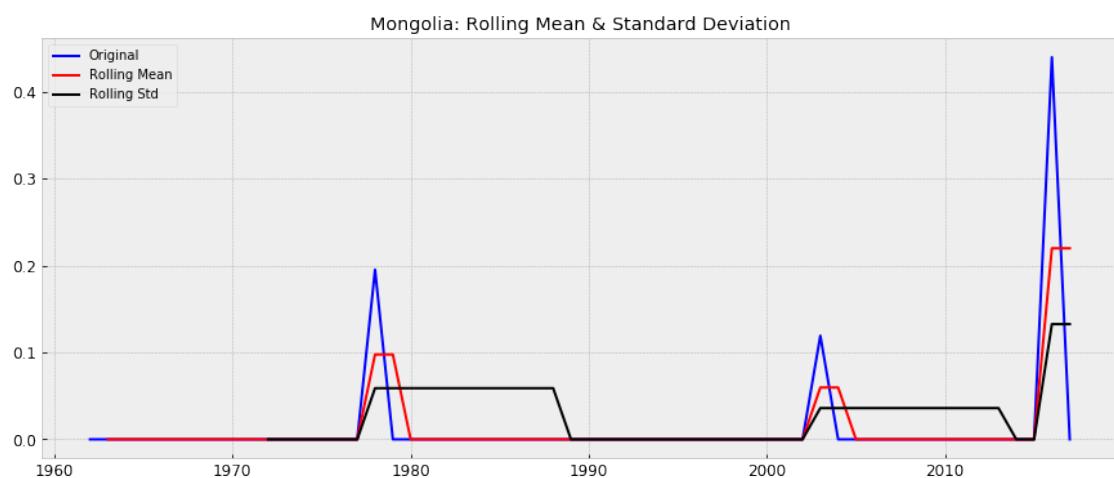
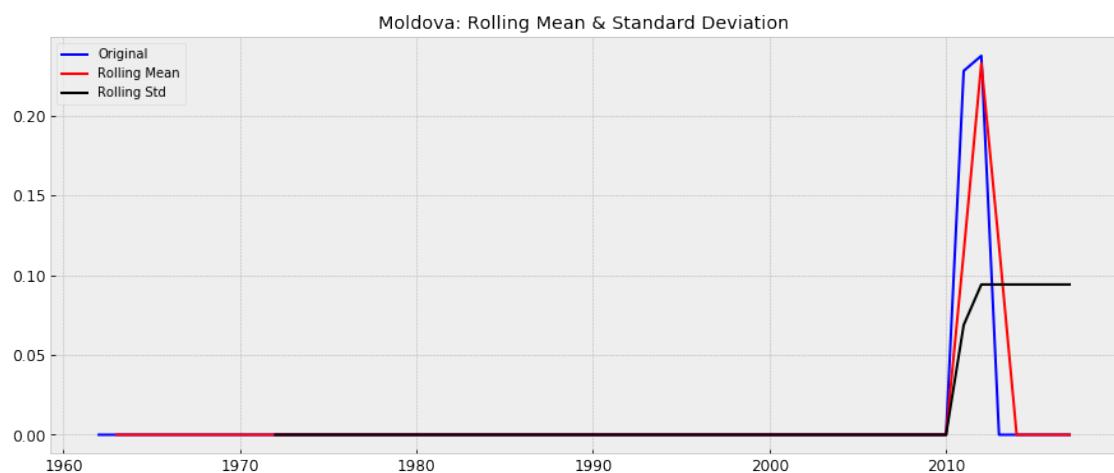
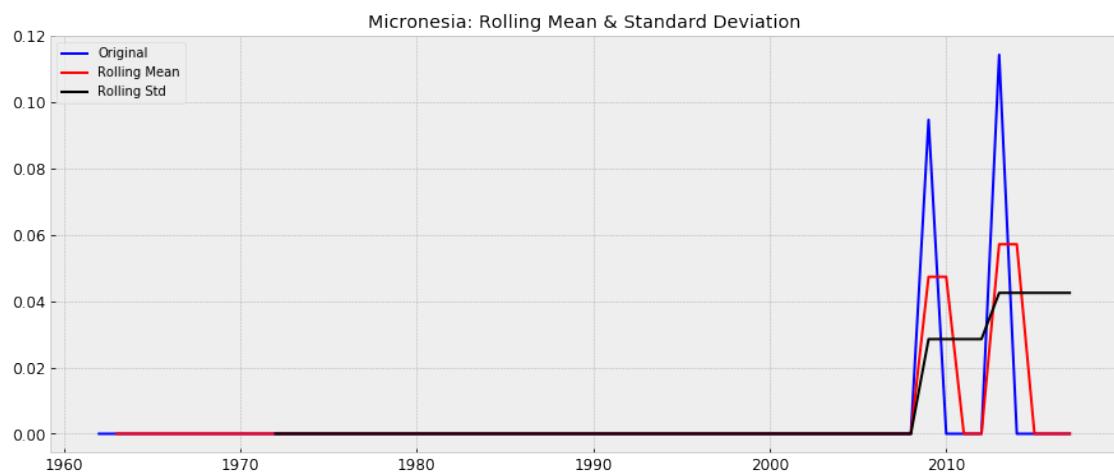


Mayotte: Rolling Mean & Standard Deviation

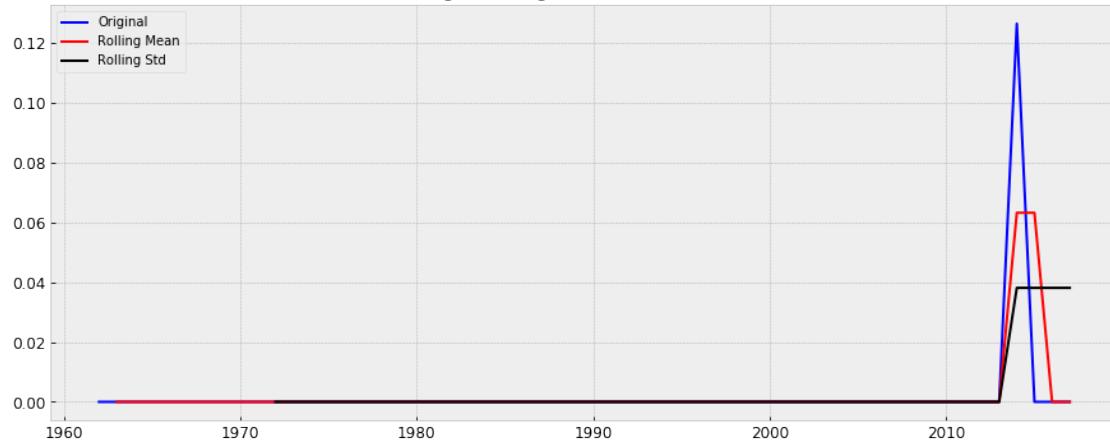


Mexico: Rolling Mean & Standard Deviation

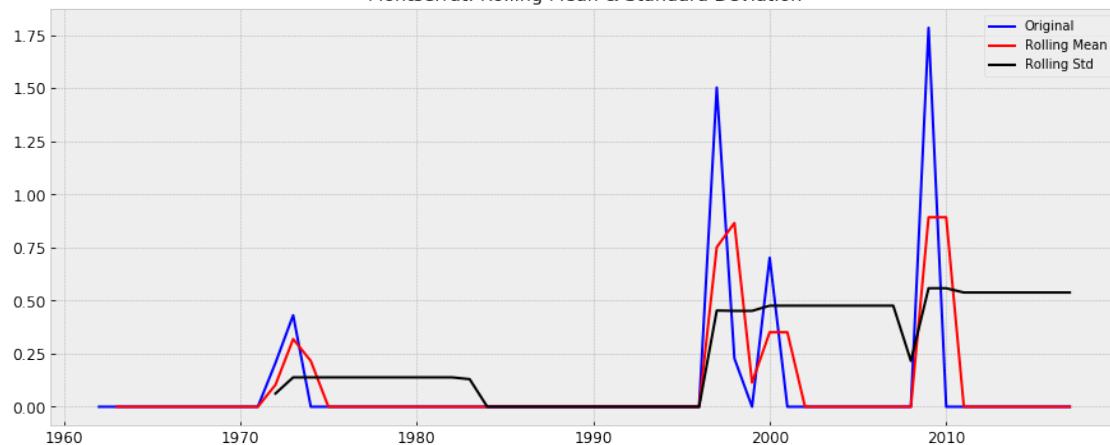




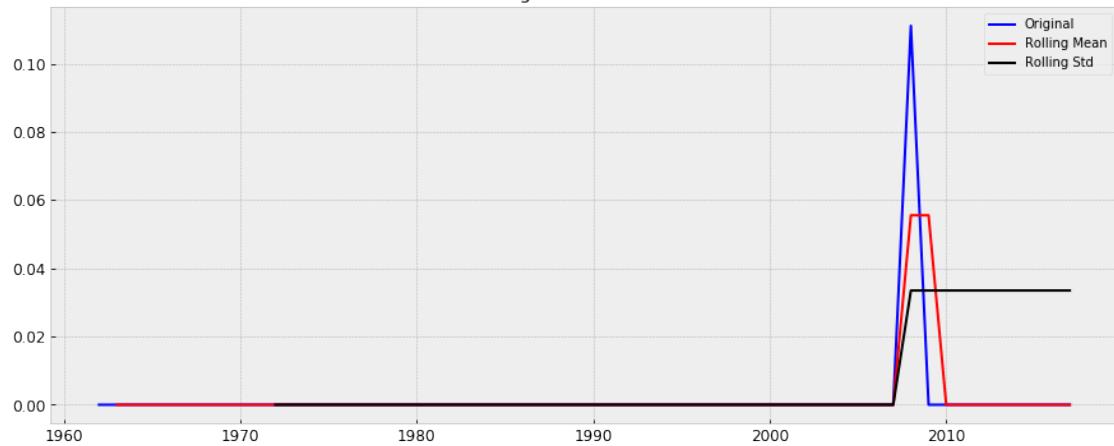
Montenegro: Rolling Mean & Standard Deviation



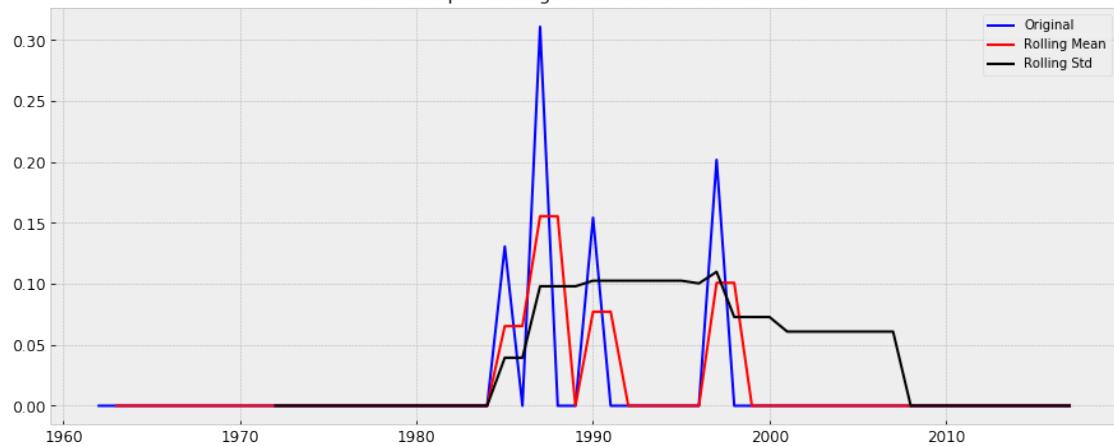
Montserrat: Rolling Mean & Standard Deviation



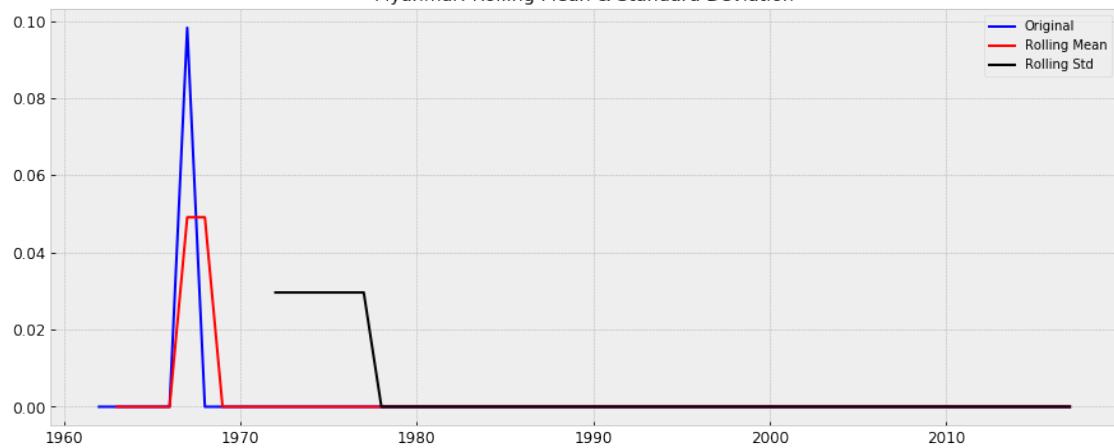
Morocco: Rolling Mean & Standard Deviation



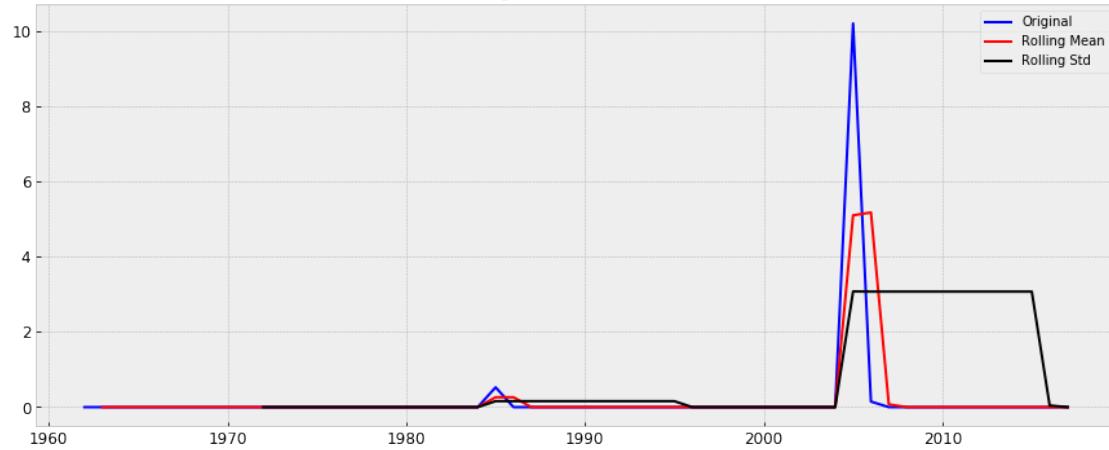
Mozambique: Rolling Mean & Standard Deviation



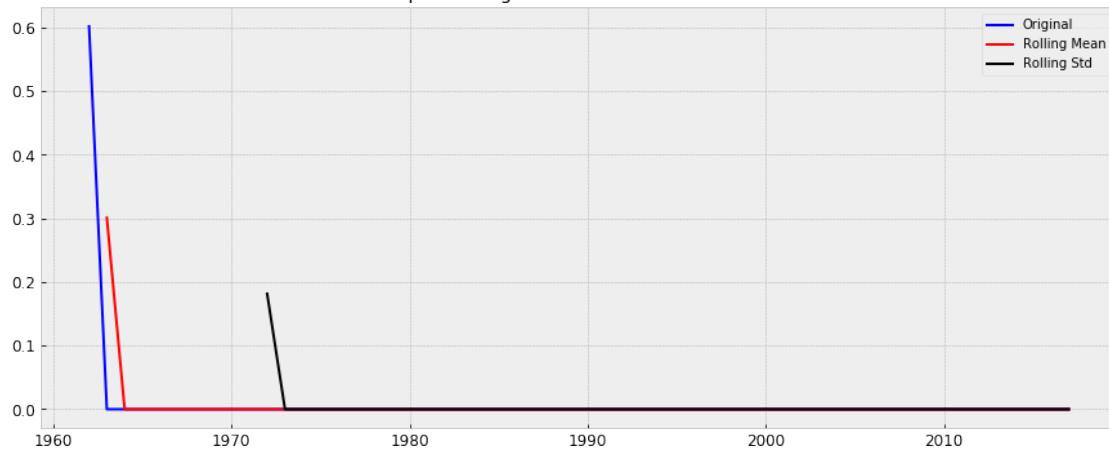
Myanmar: Rolling Mean & Standard Deviation



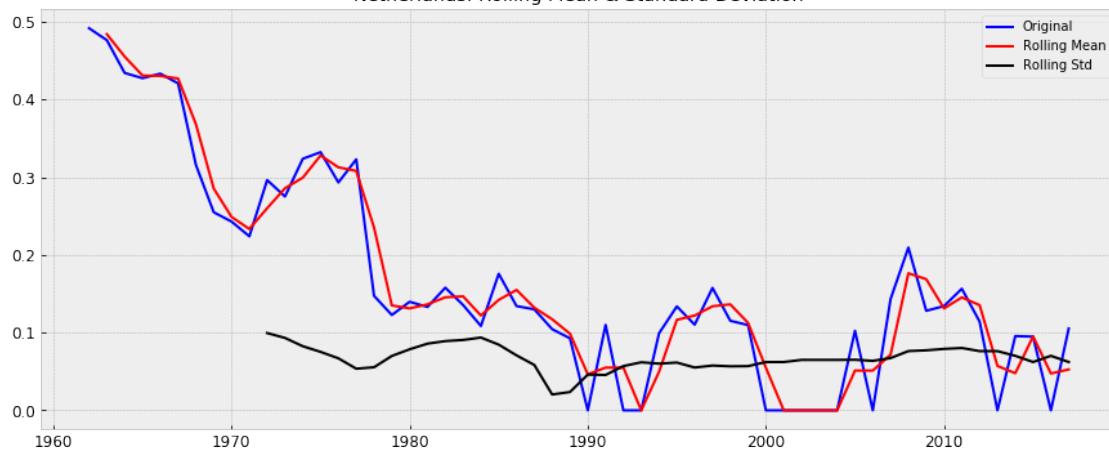
Nauru: Rolling Mean & Standard Deviation



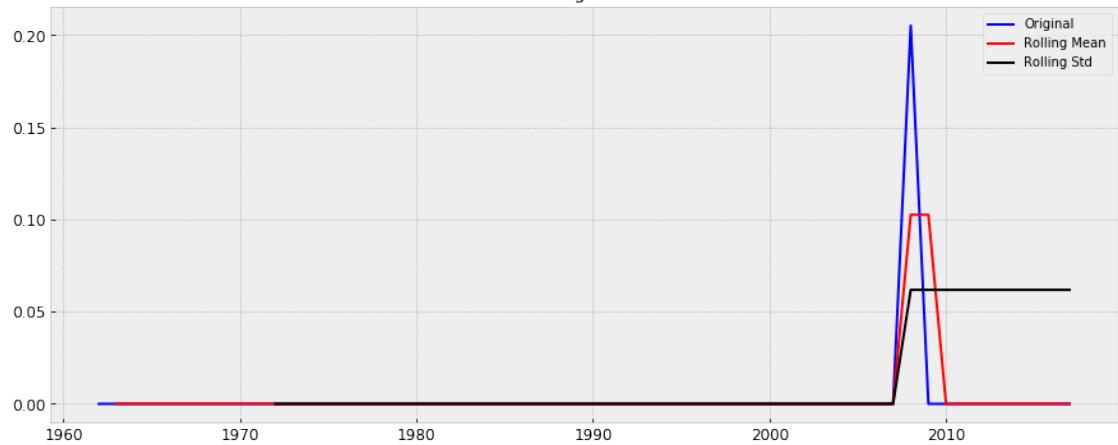
Nepal: Rolling Mean & Standard Deviation



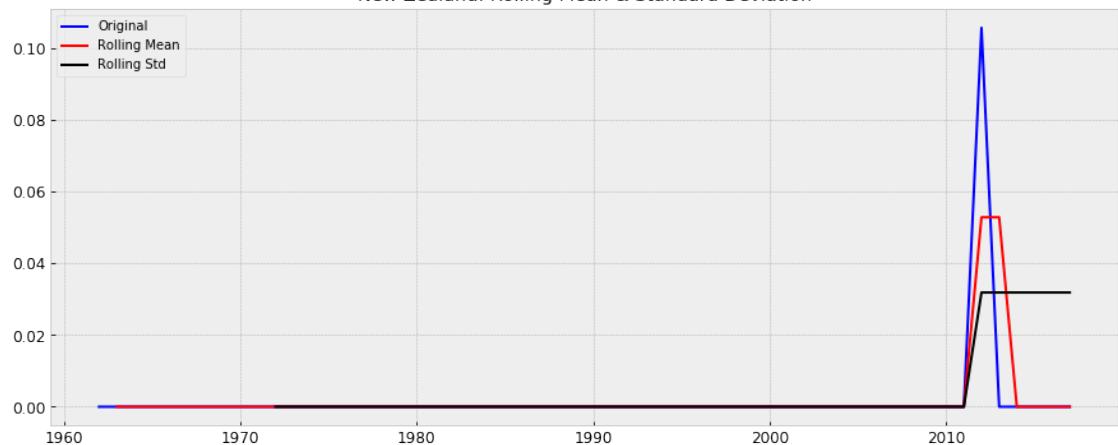
Netherlands: Rolling Mean & Standard Deviation



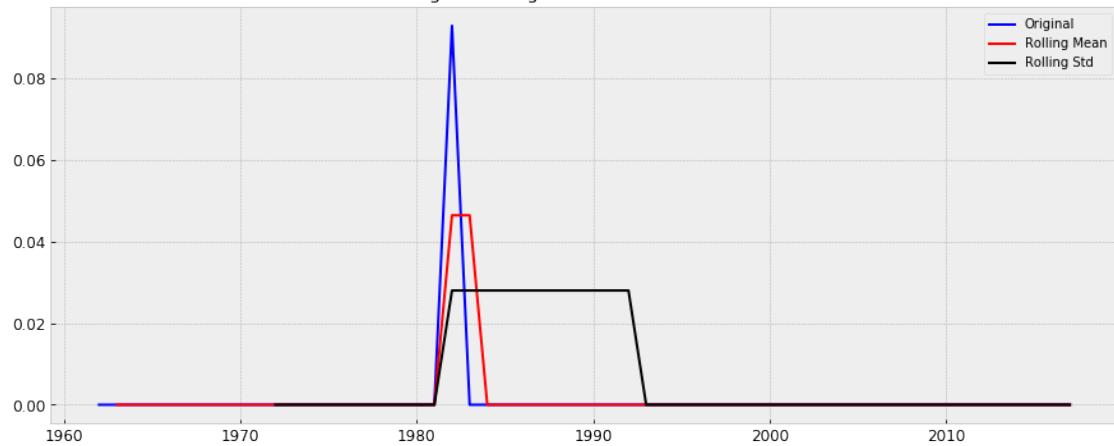
Netherlands Antilles: Rolling Mean & Standard Deviation



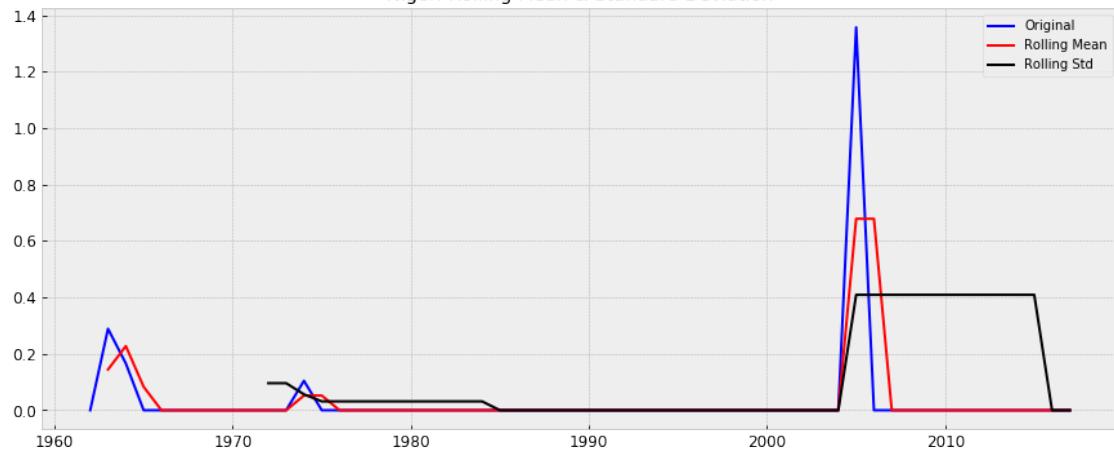
New Zealand: Rolling Mean & Standard Deviation



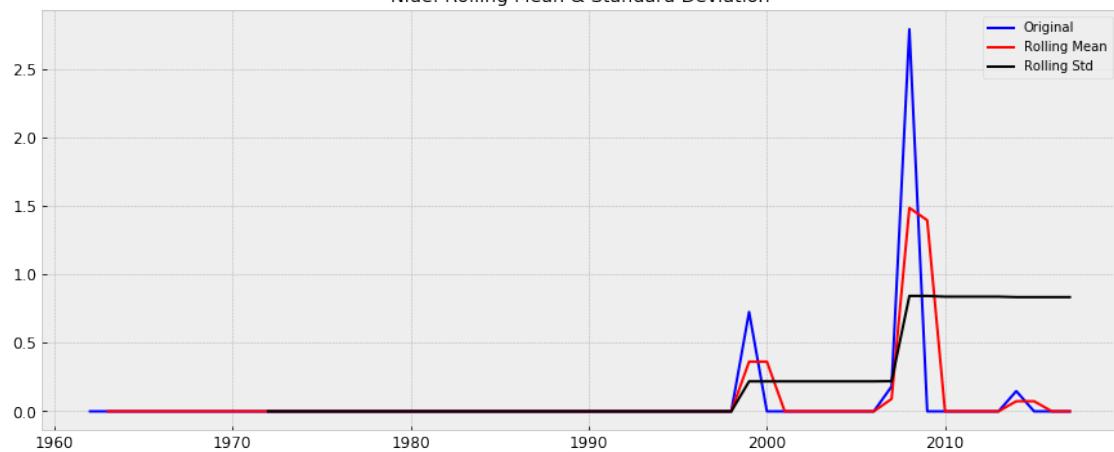
Nicaragua: Rolling Mean & Standard Deviation

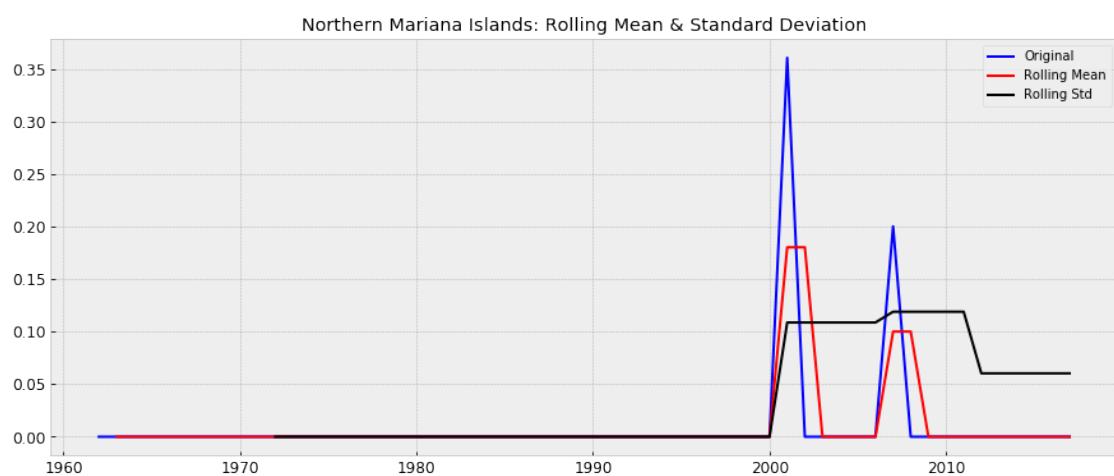
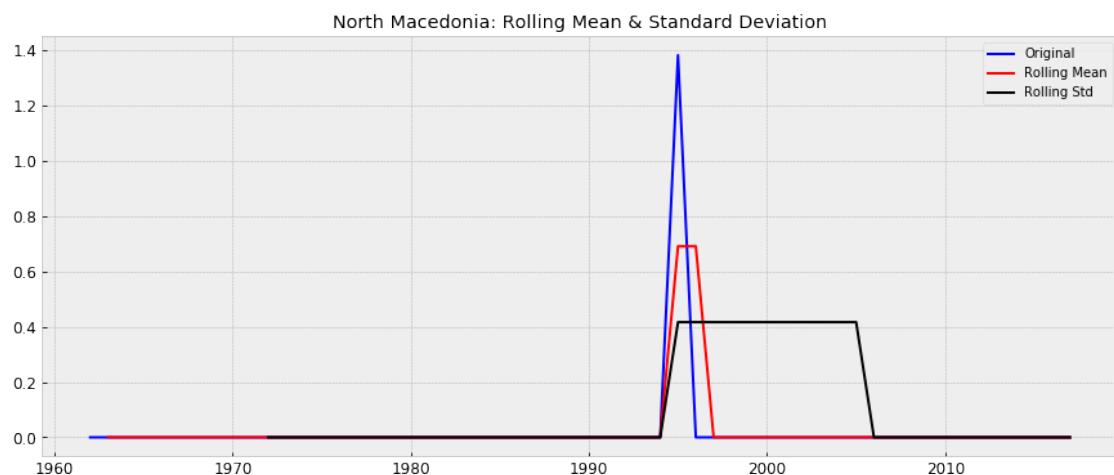
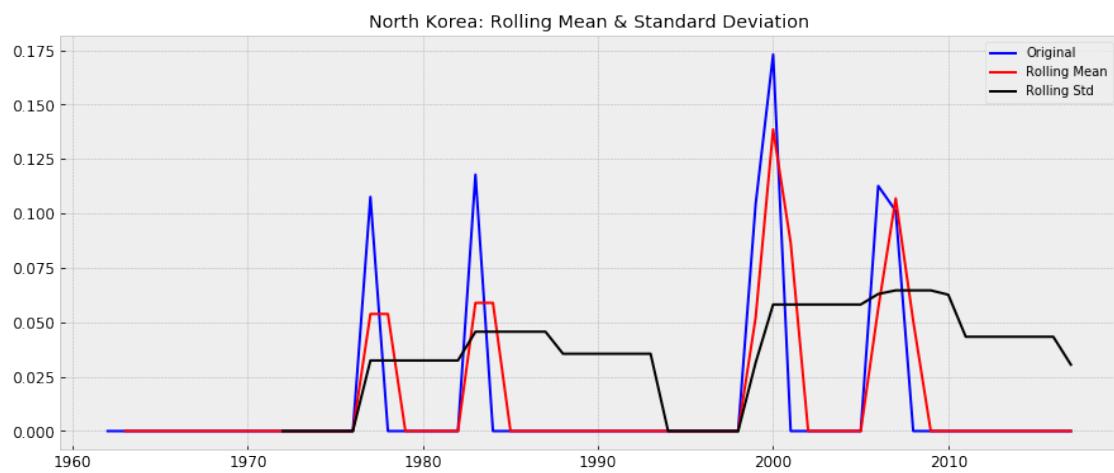


Niger: Rolling Mean & Standard Deviation

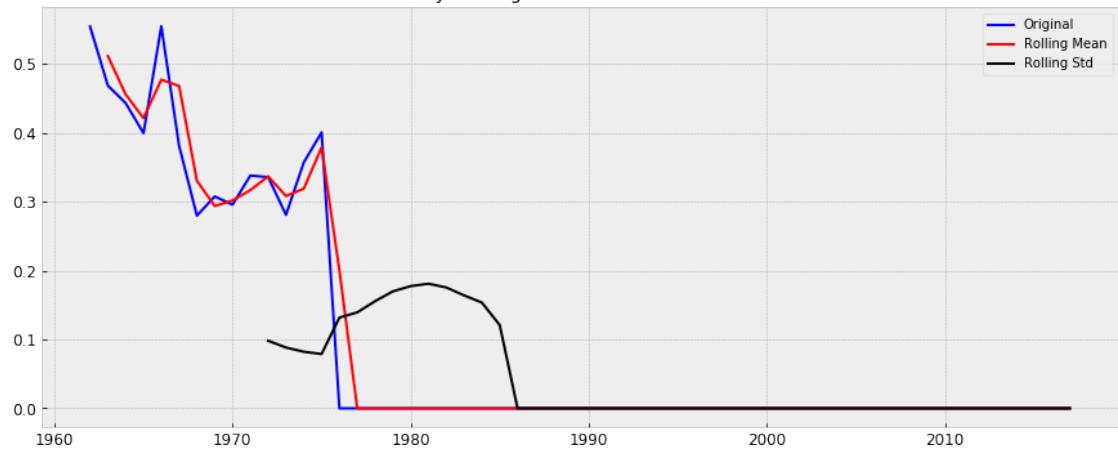


Niue: Rolling Mean & Standard Deviation

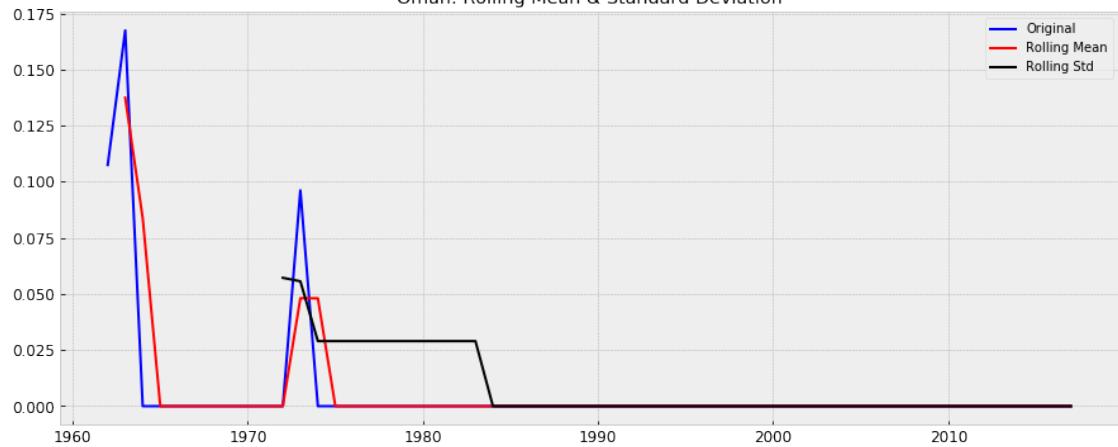




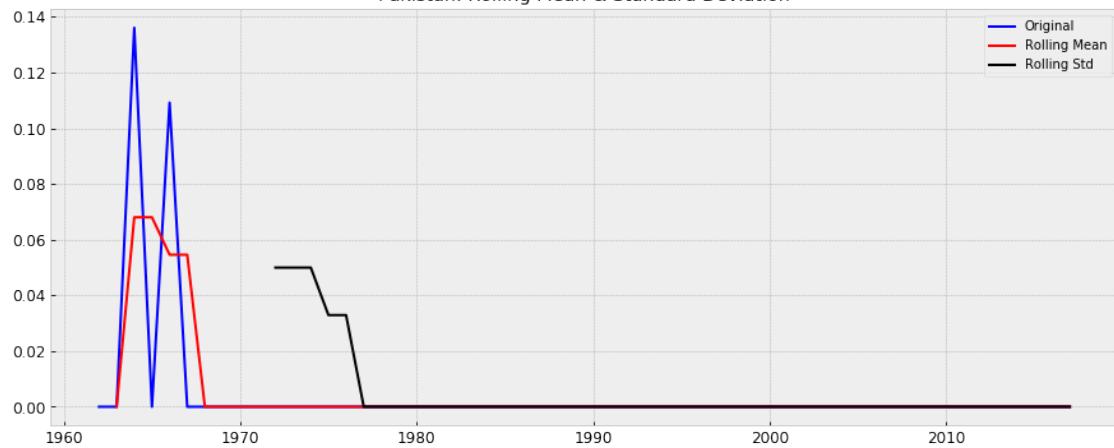
Norway: Rolling Mean & Standard Deviation



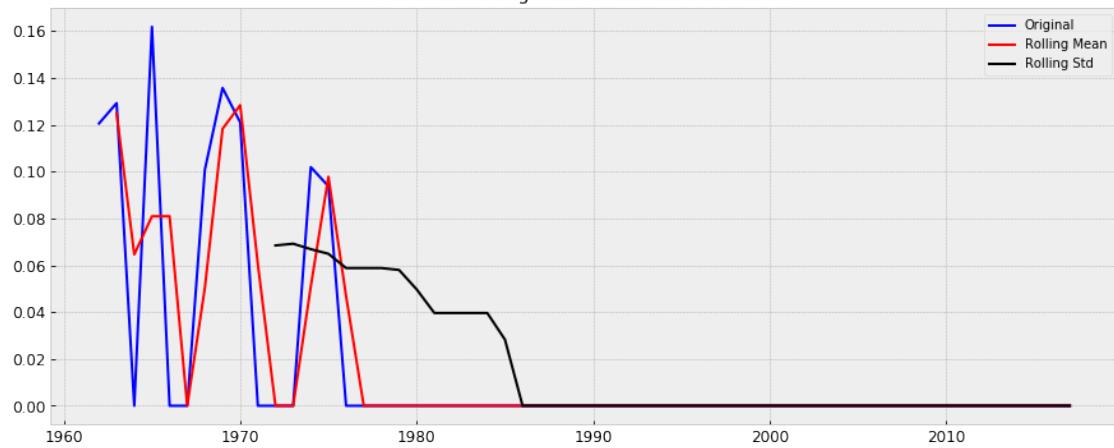
Oman: Rolling Mean & Standard Deviation



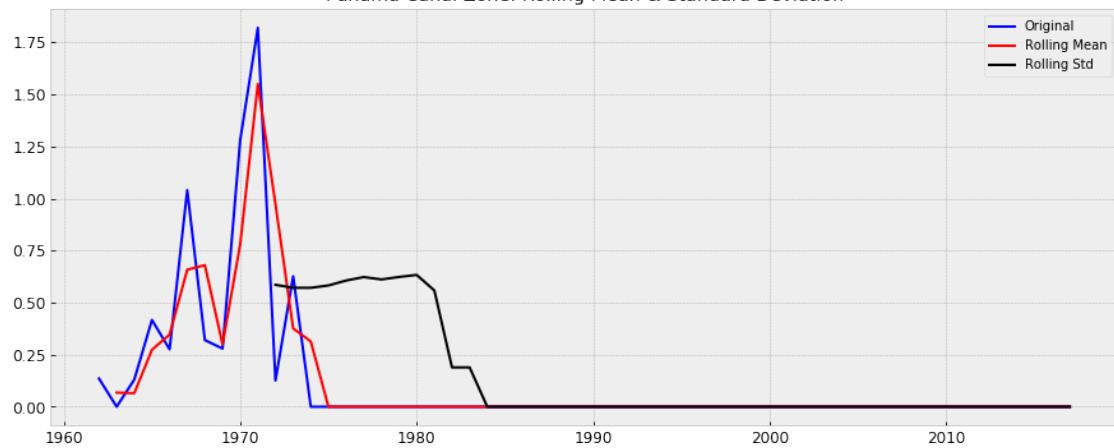
Pakistan: Rolling Mean & Standard Deviation



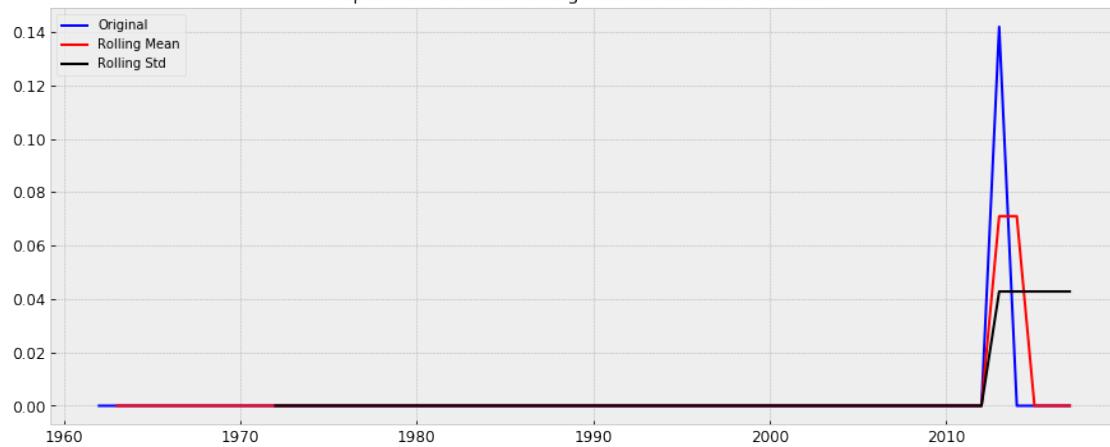
Panama: Rolling Mean & Standard Deviation



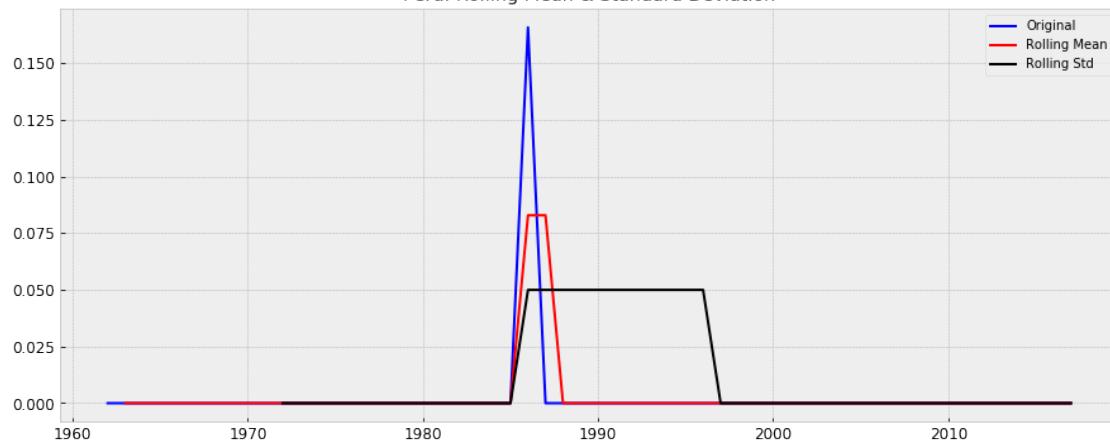
Panama Canal Zone: Rolling Mean & Standard Deviation



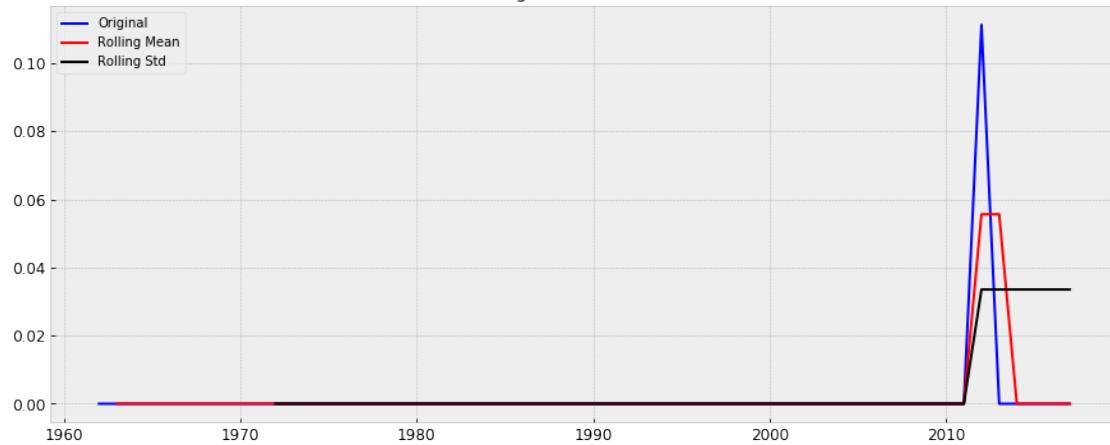
Papua New Guinea: Rolling Mean & Standard Deviation



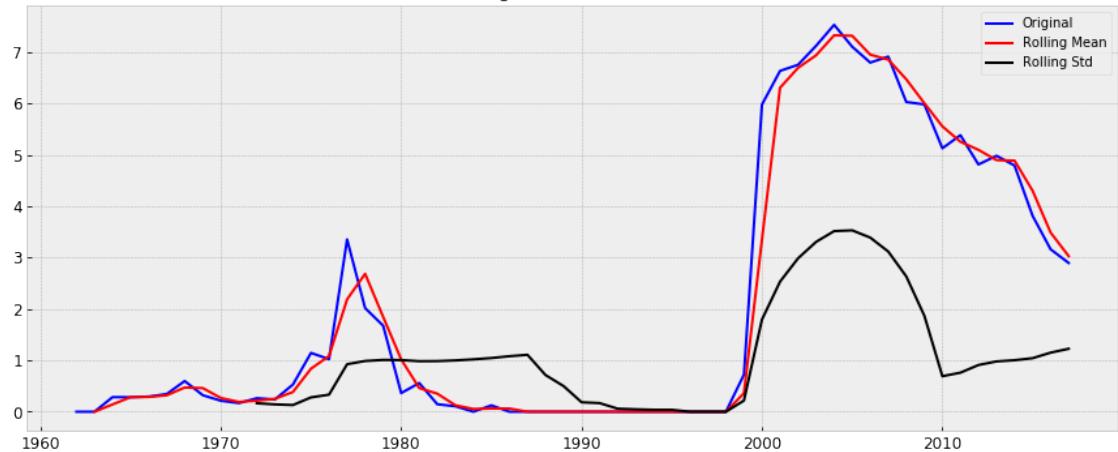
Peru: Rolling Mean & Standard Deviation



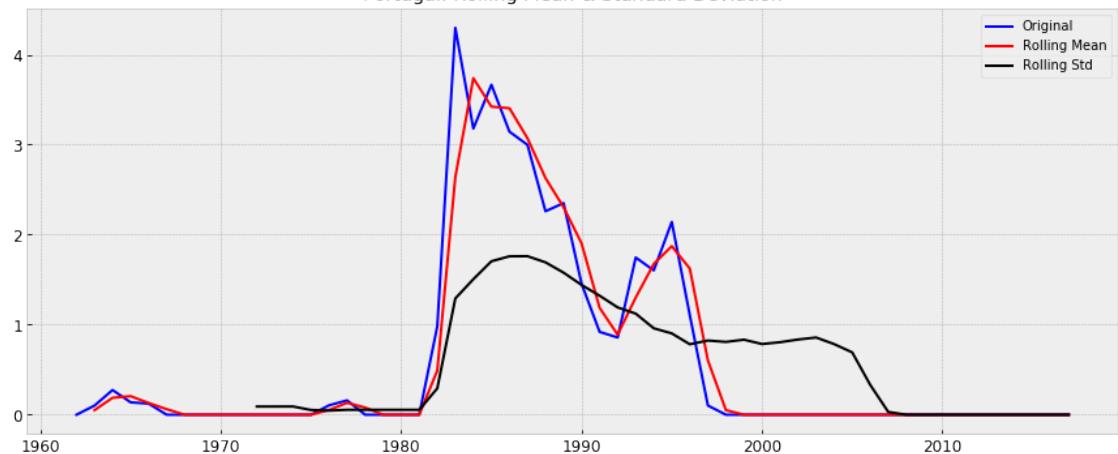
Pitcairn: Rolling Mean & Standard Deviation



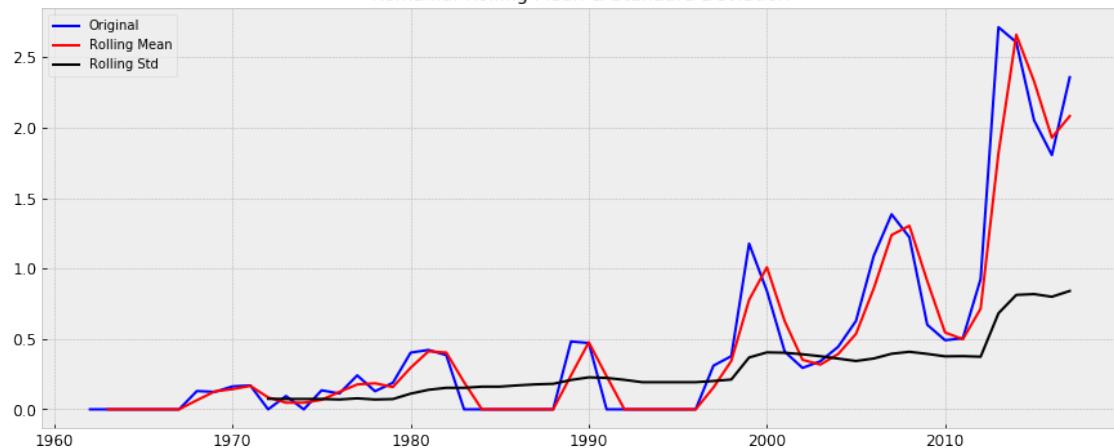
Poland: Rolling Mean & Standard Deviation



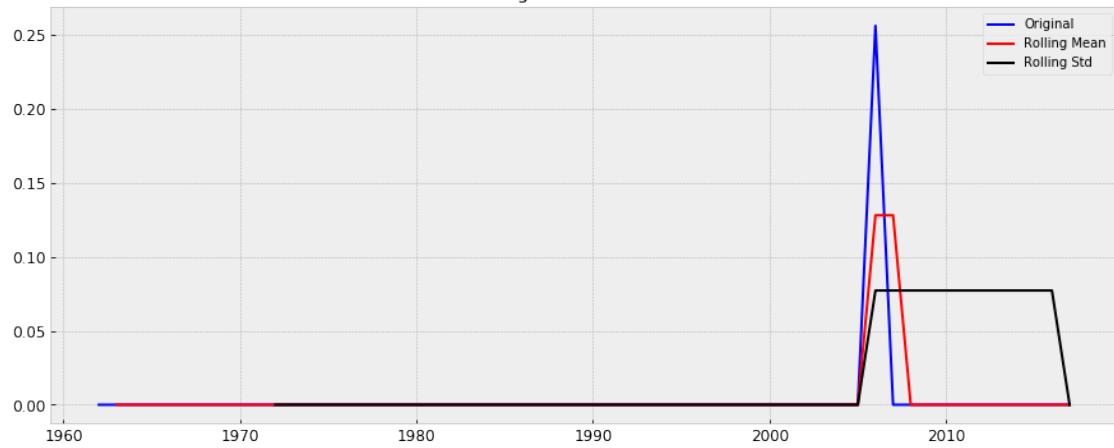
Portugal: Rolling Mean & Standard Deviation



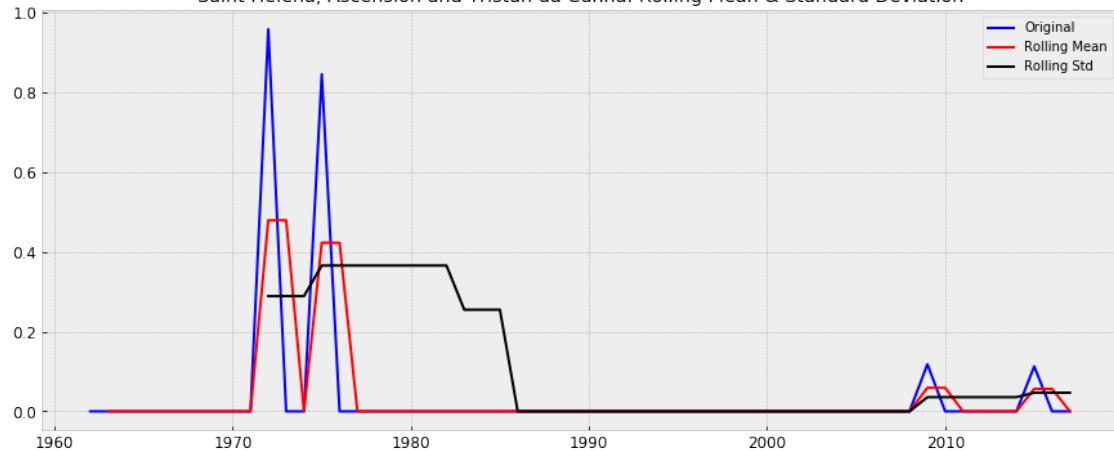
Romania: Rolling Mean & Standard Deviation



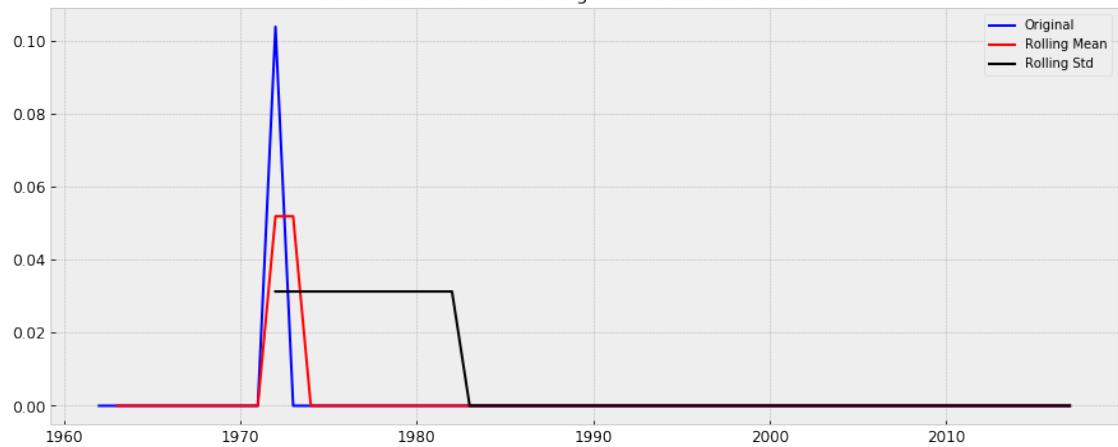
Rwanda: Rolling Mean & Standard Deviation



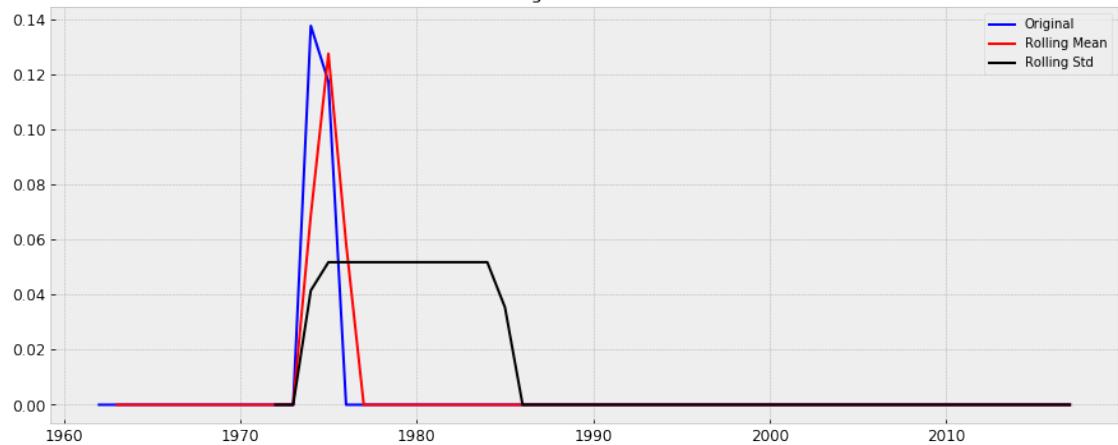
Saint Helena, Ascension and Tristan da Cunha: Rolling Mean & Standard Deviation



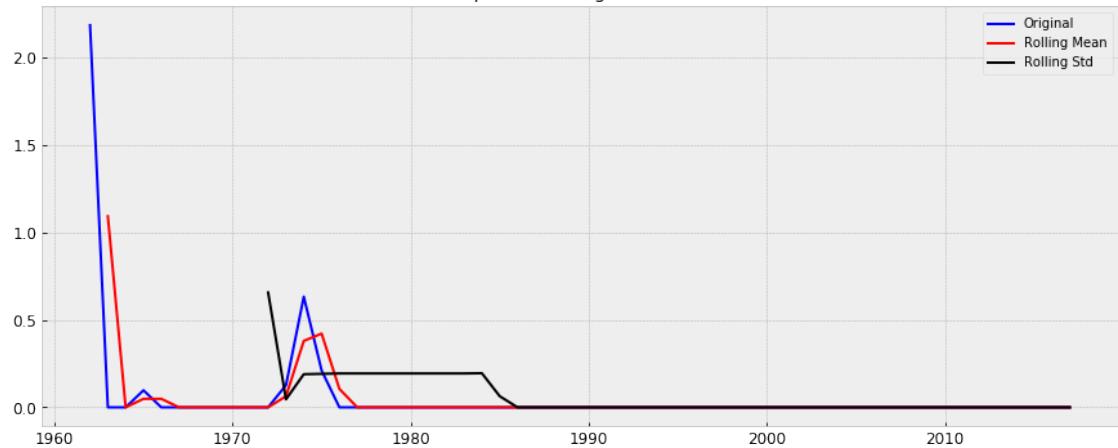
Saint Kitts and Nevis: Rolling Mean & Standard Deviation



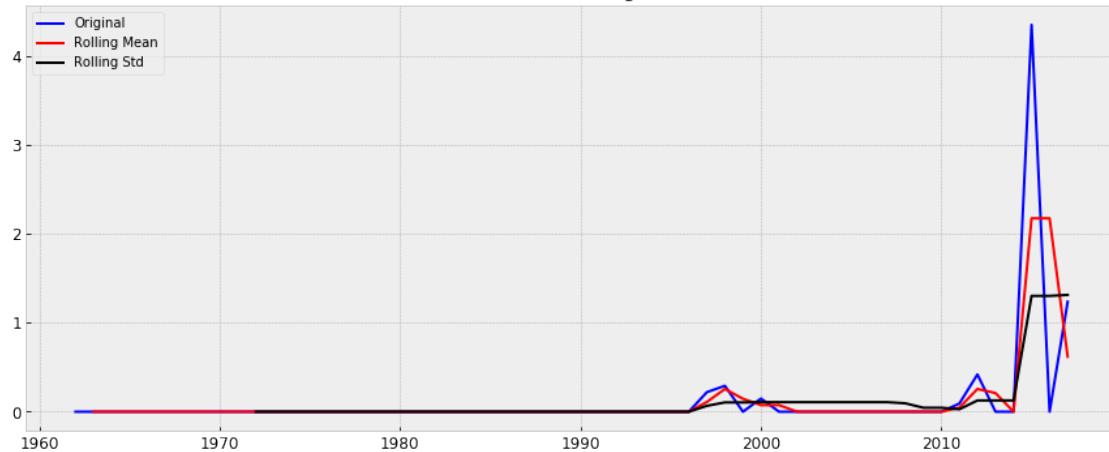
Saint Lucia: Rolling Mean & Standard Deviation



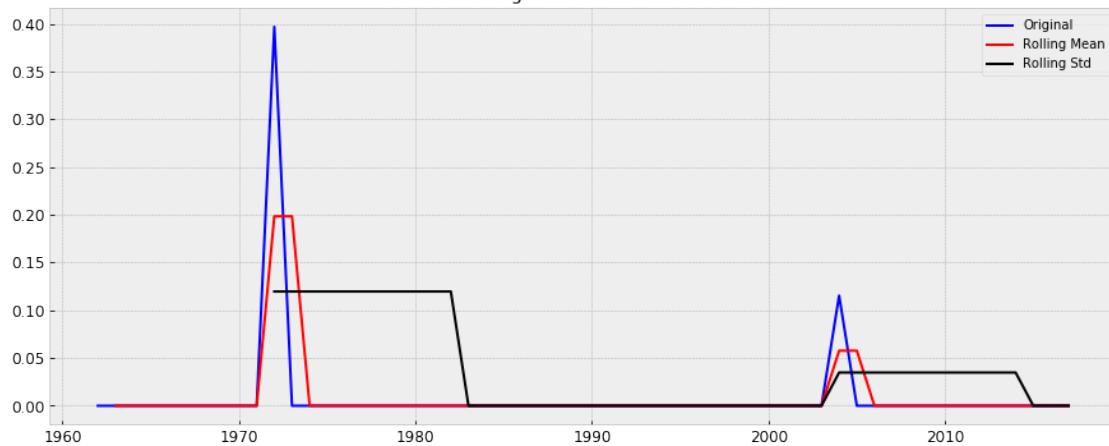
Saint Pierre and Miquelon: Rolling Mean & Standard Deviation

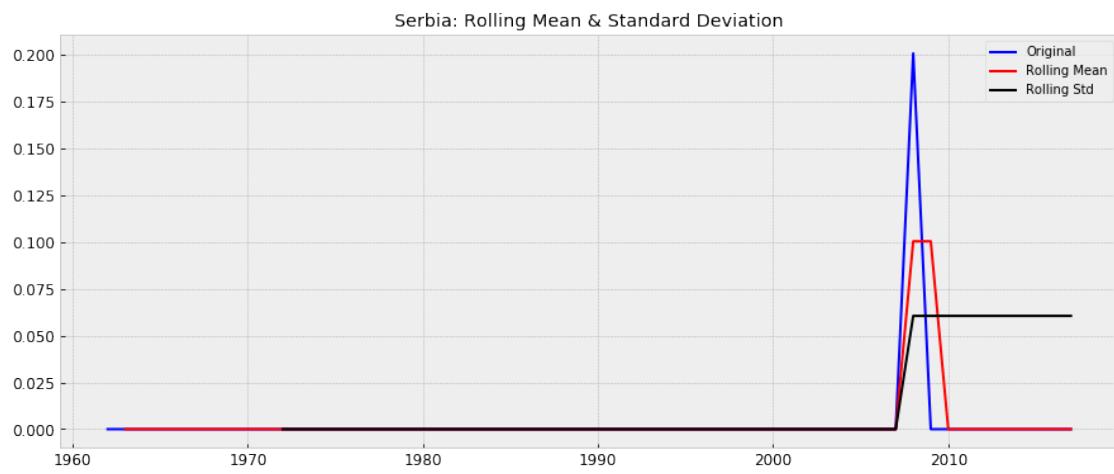
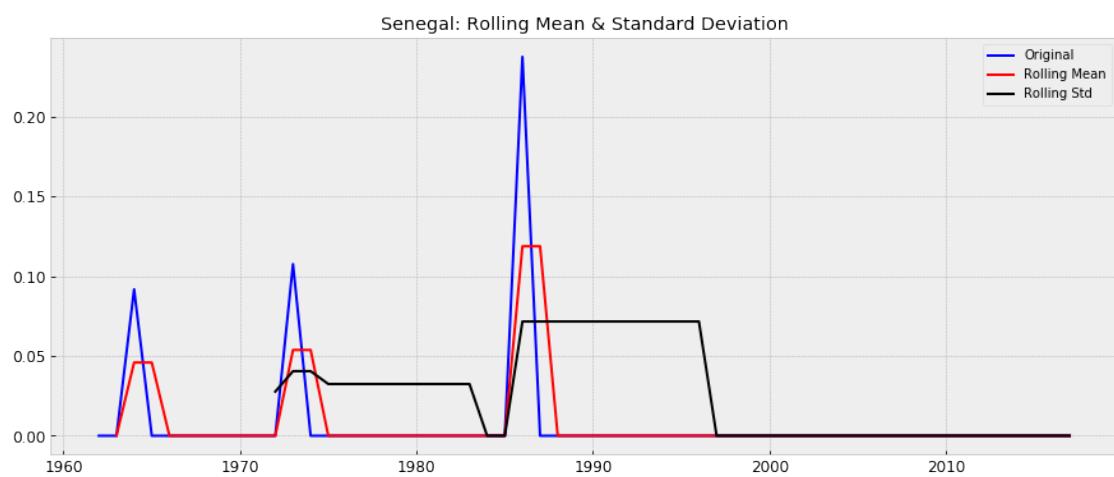
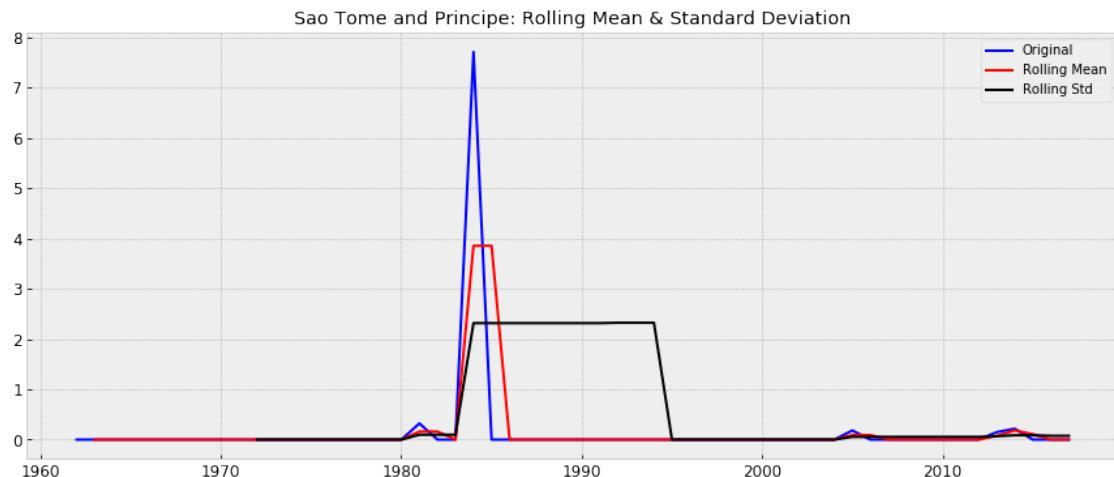


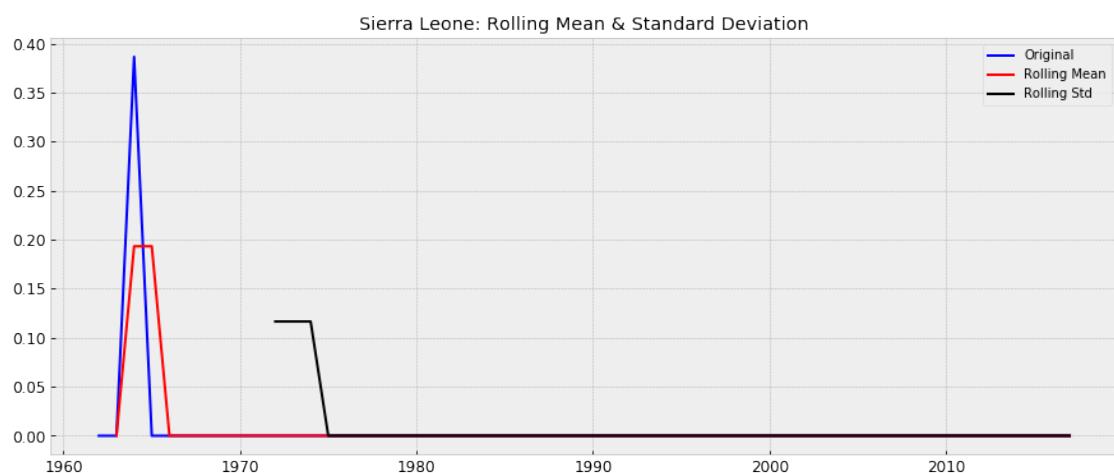
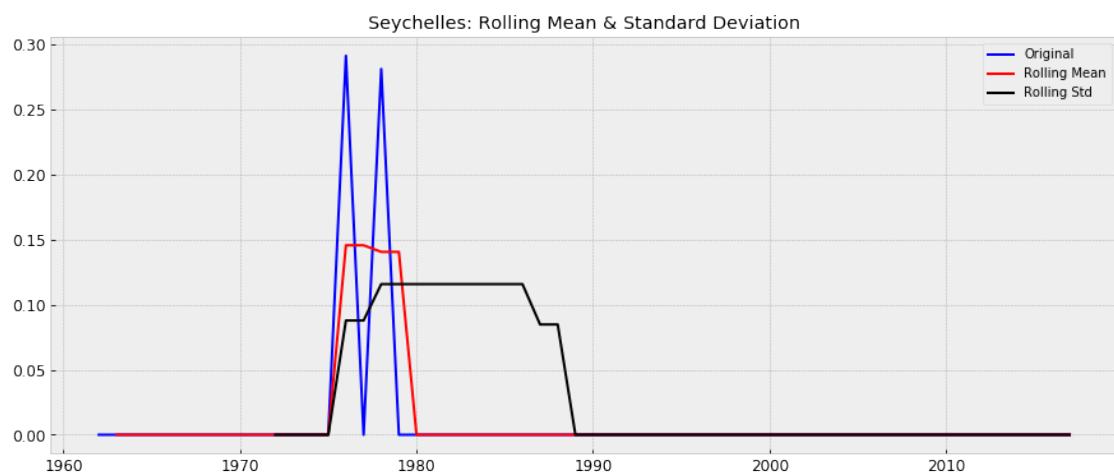
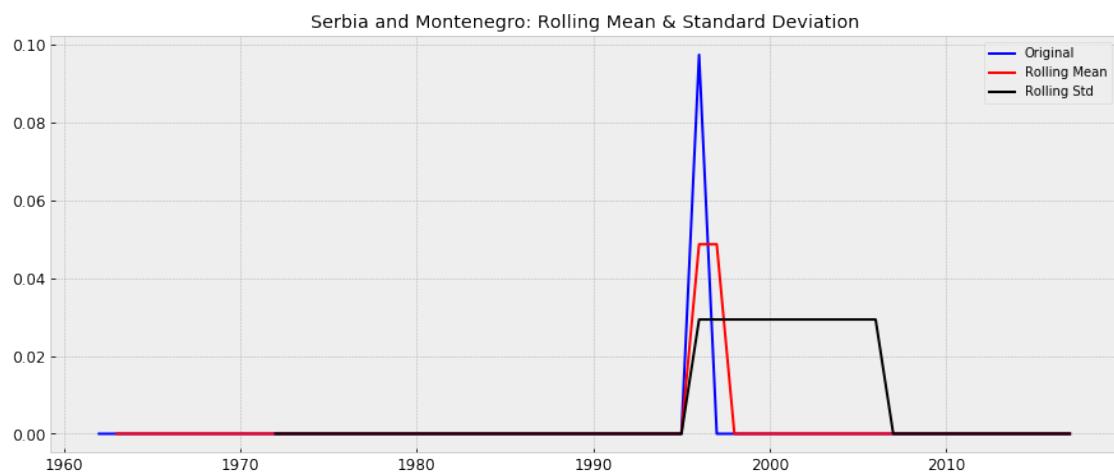
Saint Vincent and the Grenadines: Rolling Mean & Standard Deviation



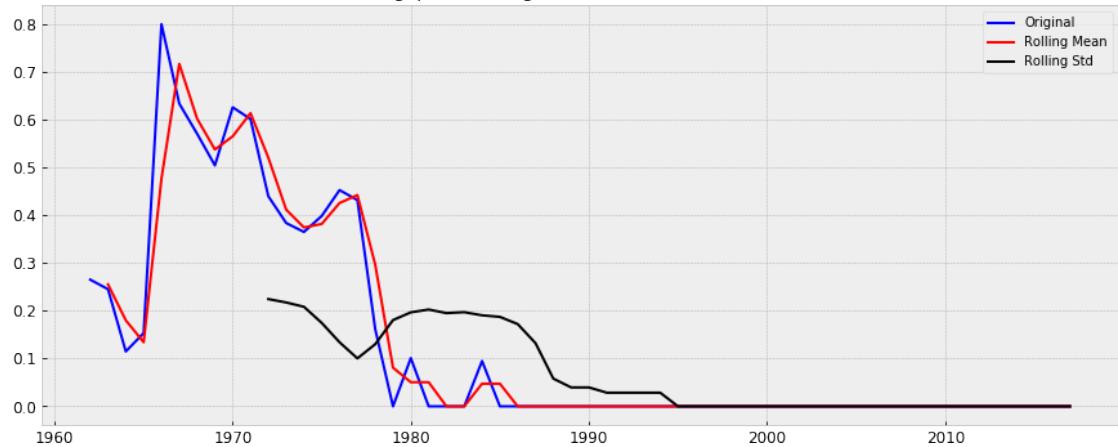
Samoa: Rolling Mean & Standard Deviation



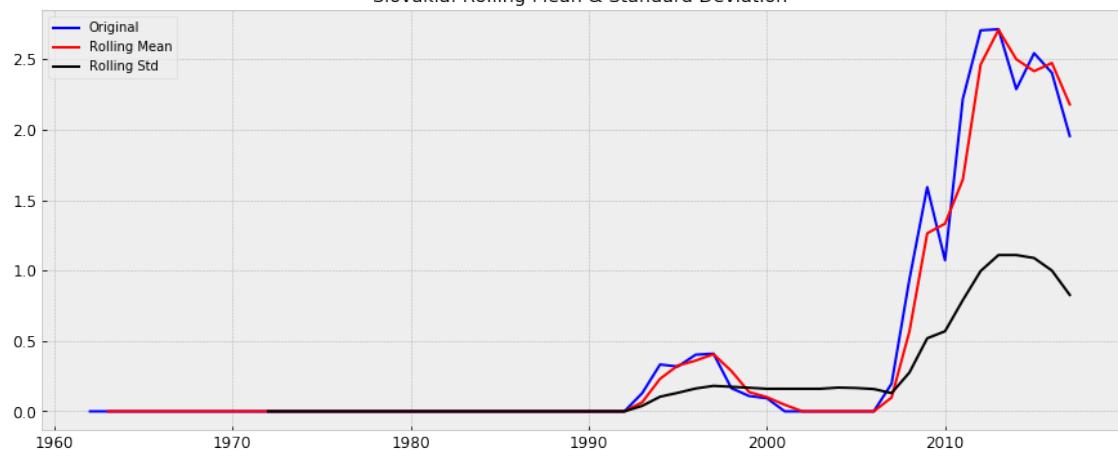




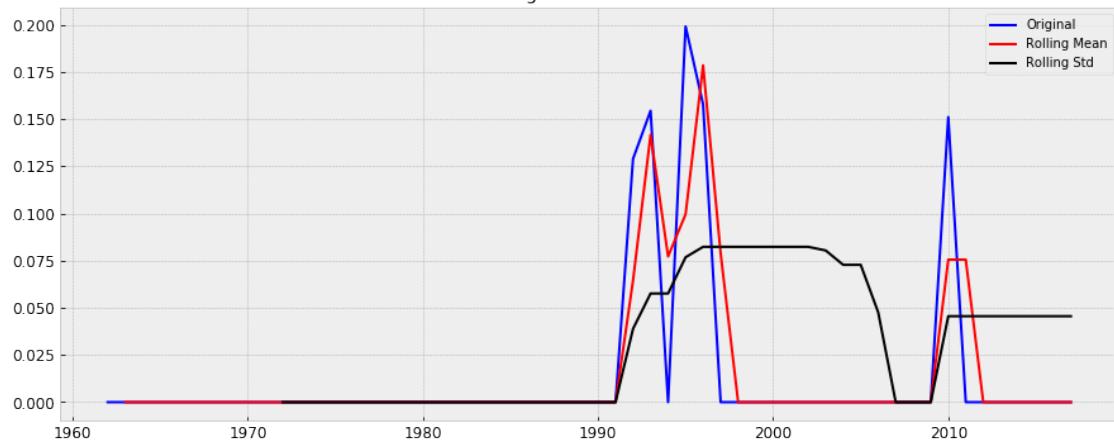
Singapore: Rolling Mean & Standard Deviation



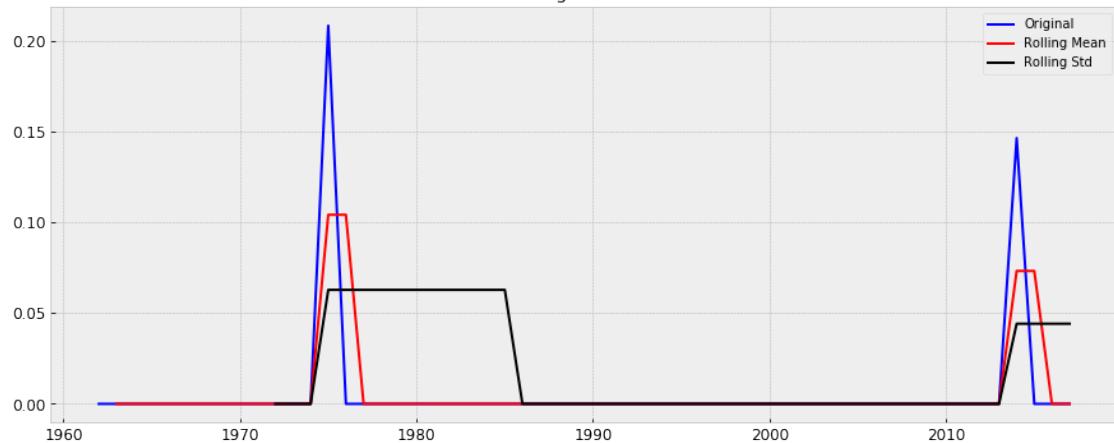
Slovakia: Rolling Mean & Standard Deviation



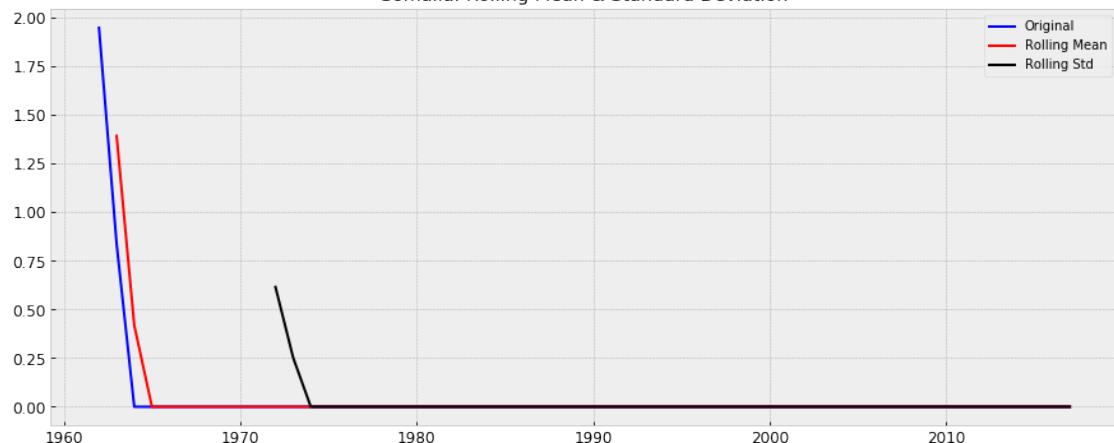
Slovenia: Rolling Mean & Standard Deviation



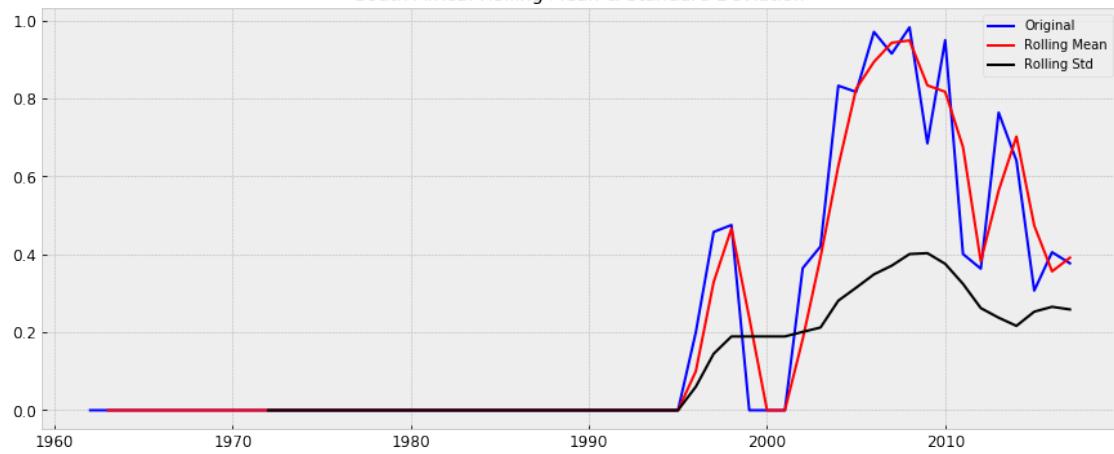
Solomon Islands: Rolling Mean & Standard Deviation



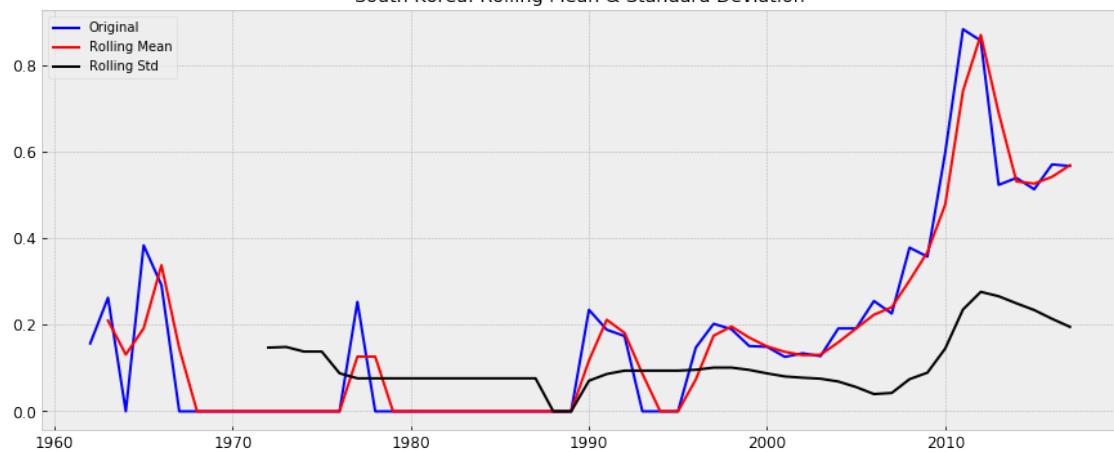
Somalia: Rolling Mean & Standard Deviation



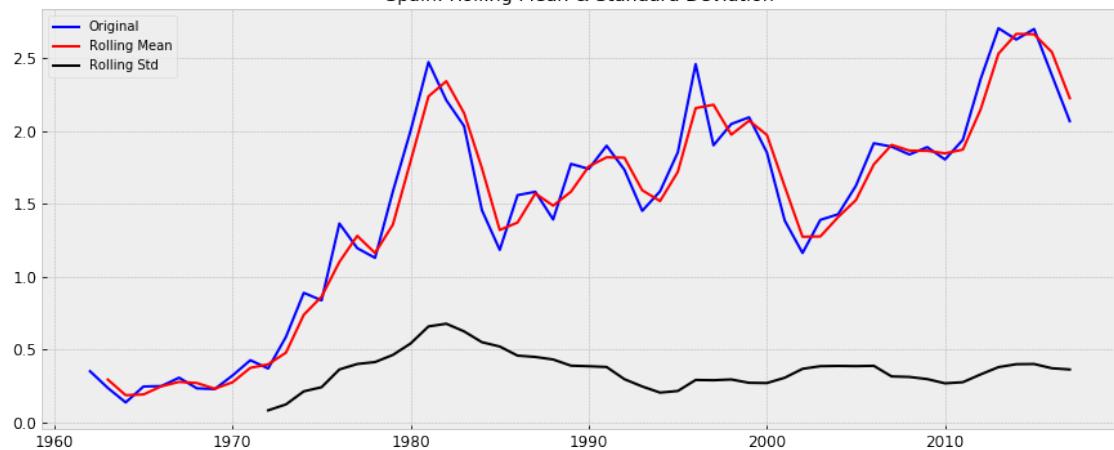
South Africa: Rolling Mean & Standard Deviation

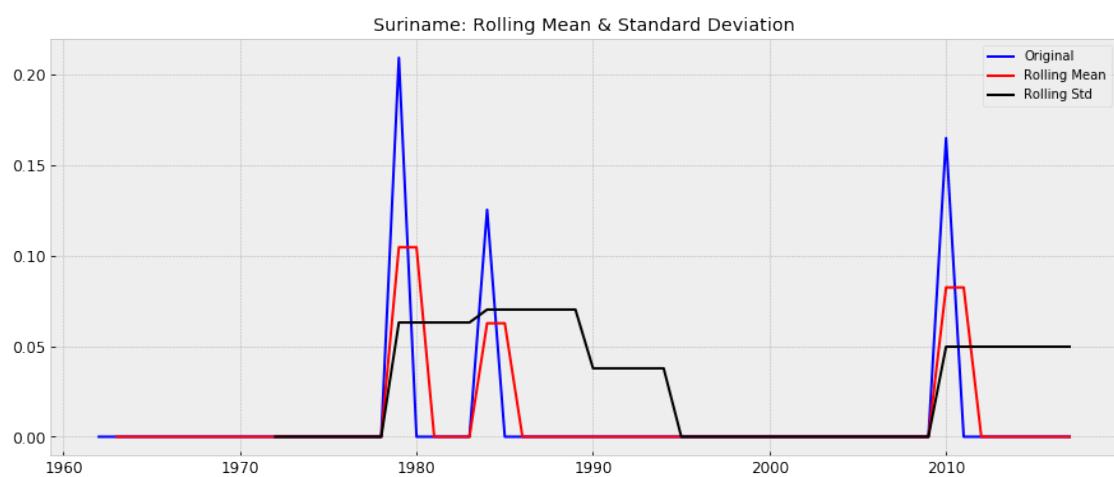
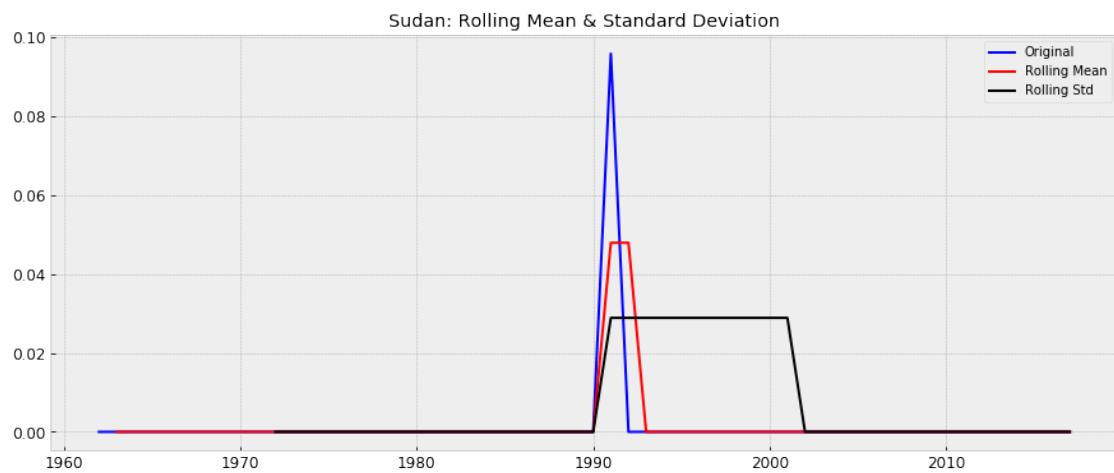


South Korea: Rolling Mean & Standard Deviation

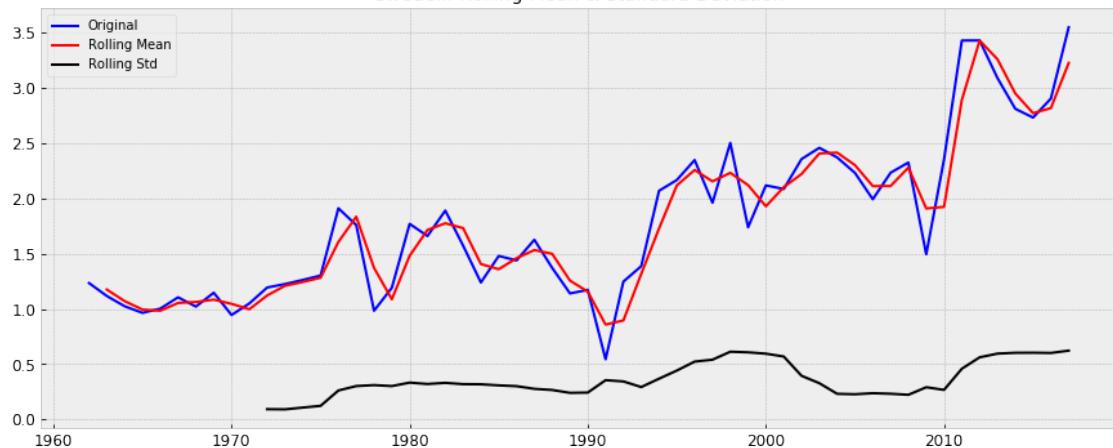


Spain: Rolling Mean & Standard Deviation

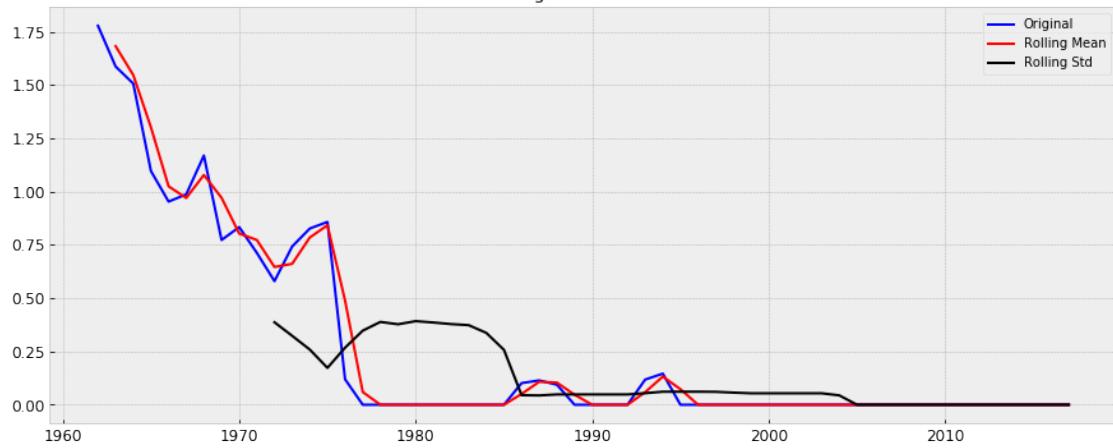




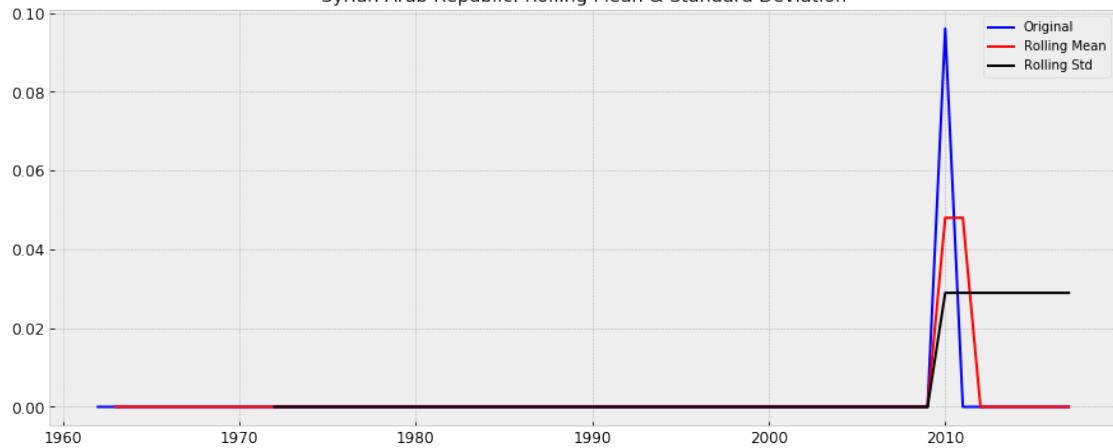
Sweden: Rolling Mean & Standard Deviation

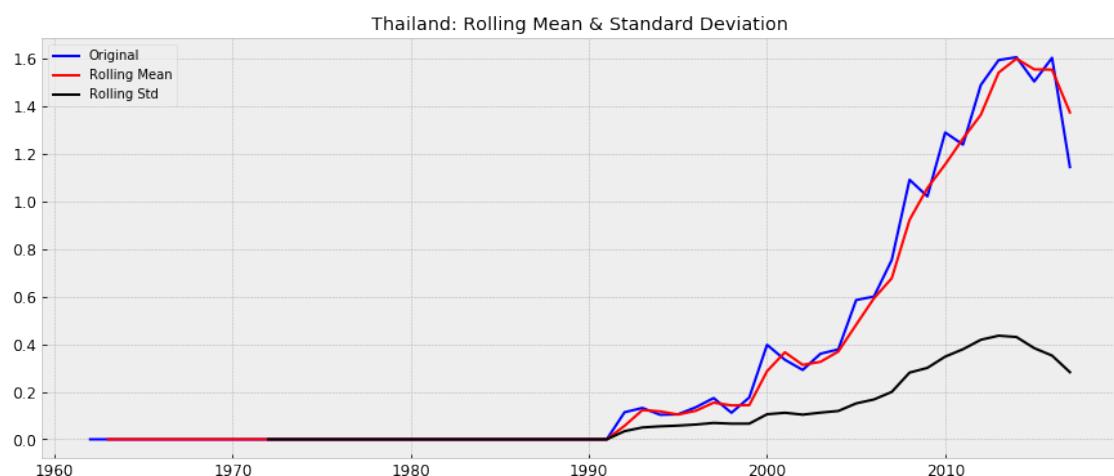
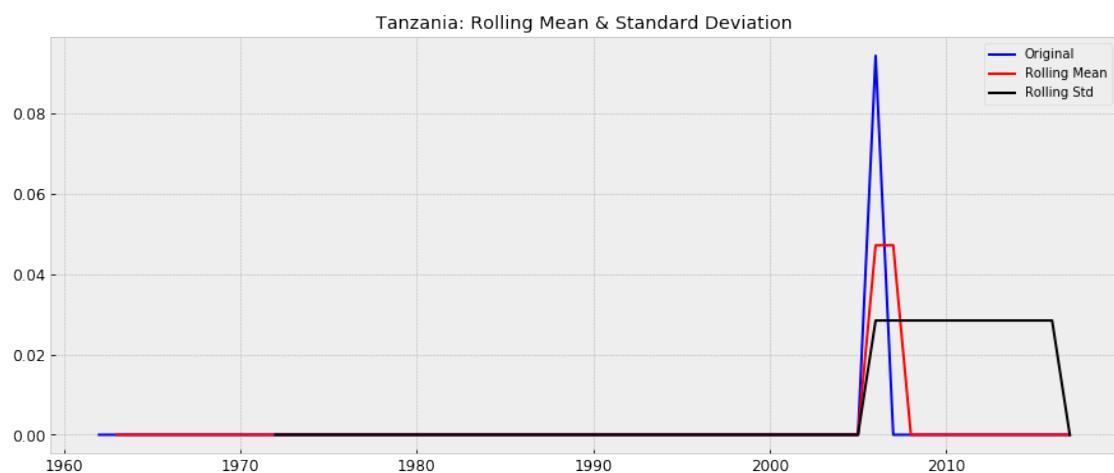
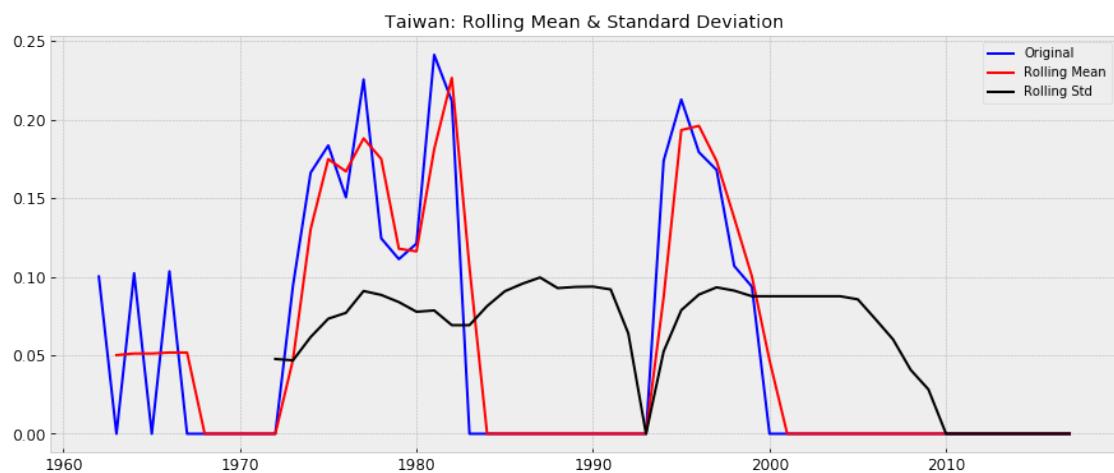


Switzerland: Rolling Mean & Standard Deviation

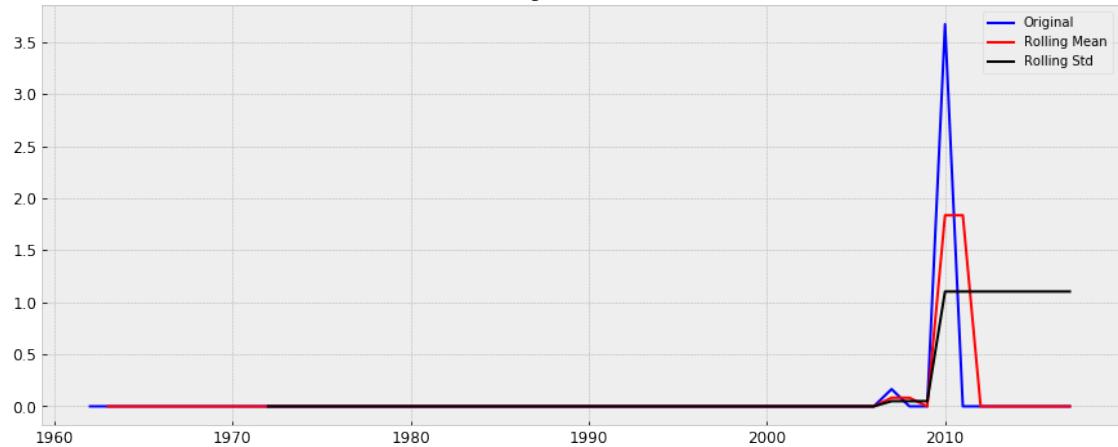


Syrian Arab Republic: Rolling Mean & Standard Deviation

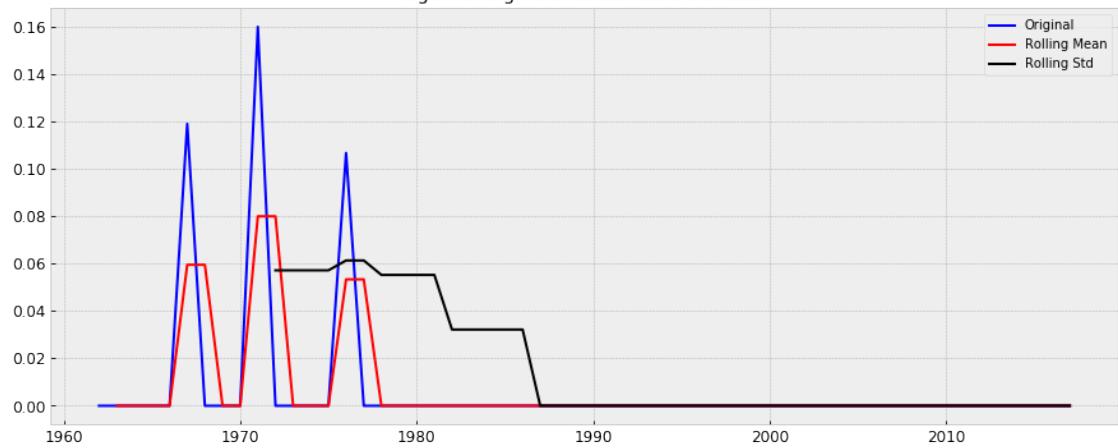




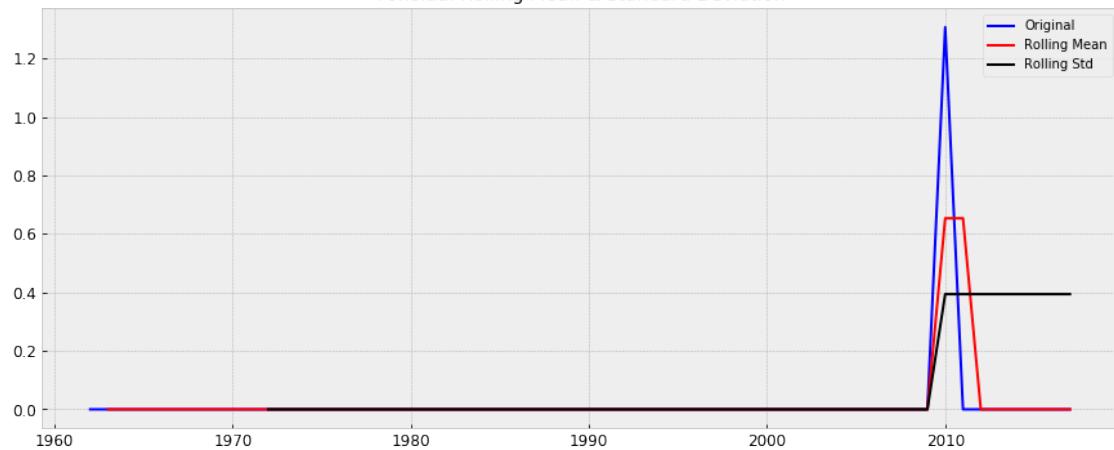
Timor-Leste: Rolling Mean & Standard Deviation



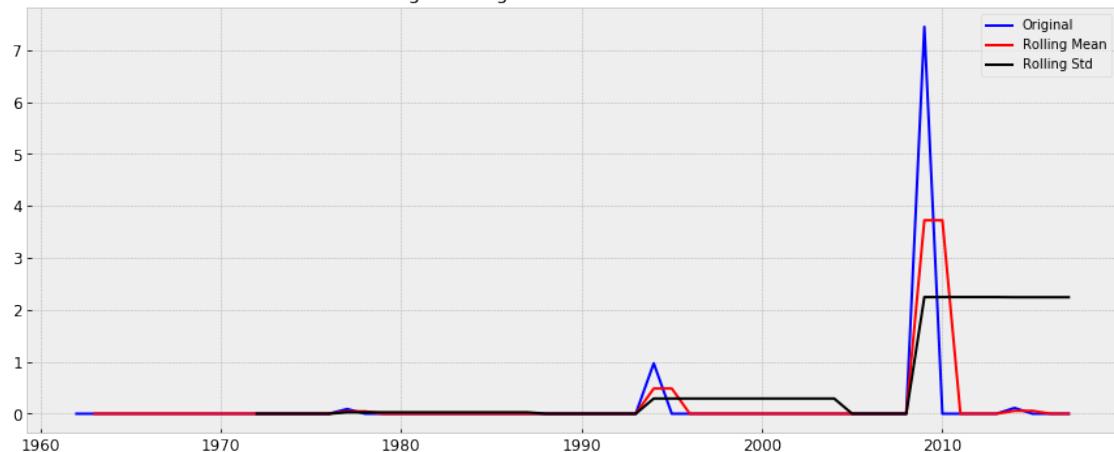
Togo: Rolling Mean & Standard Deviation



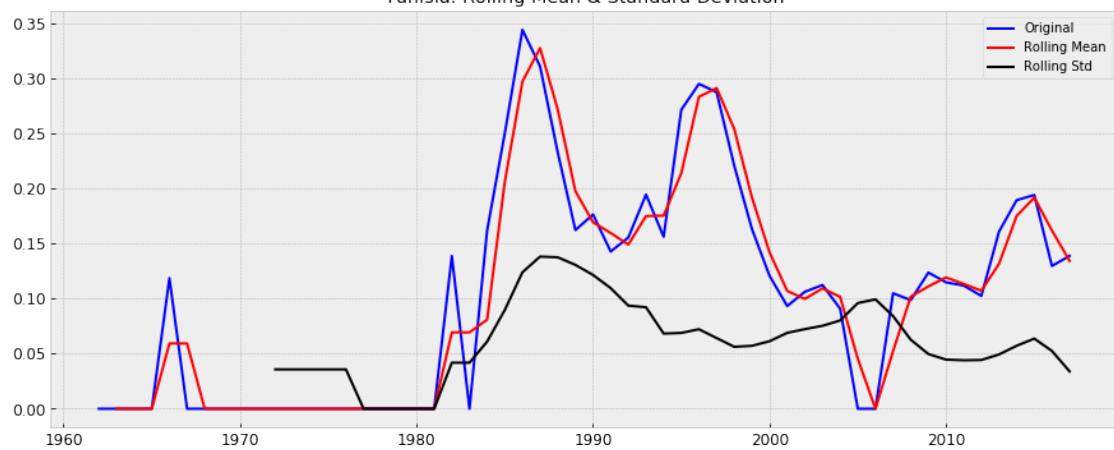
Tokelau: Rolling Mean & Standard Deviation



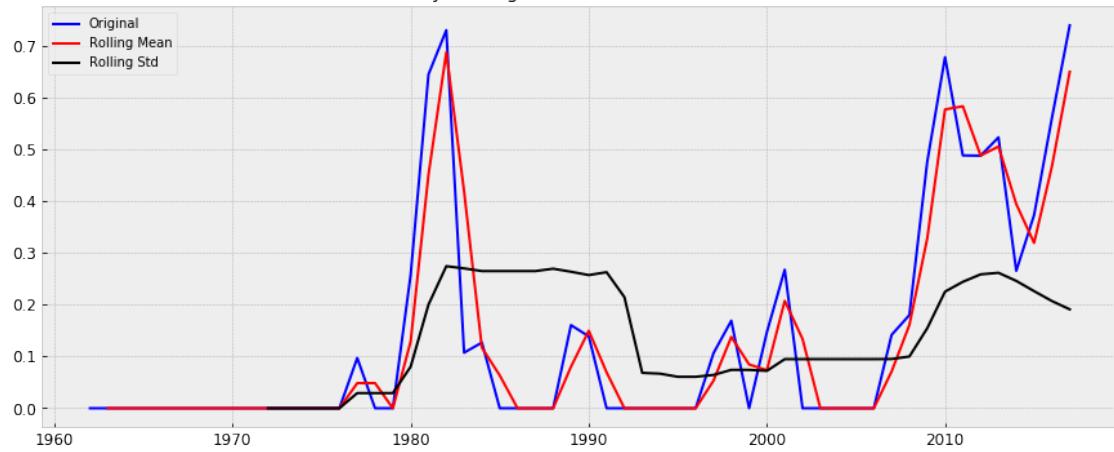
Tonga: Rolling Mean & Standard Deviation



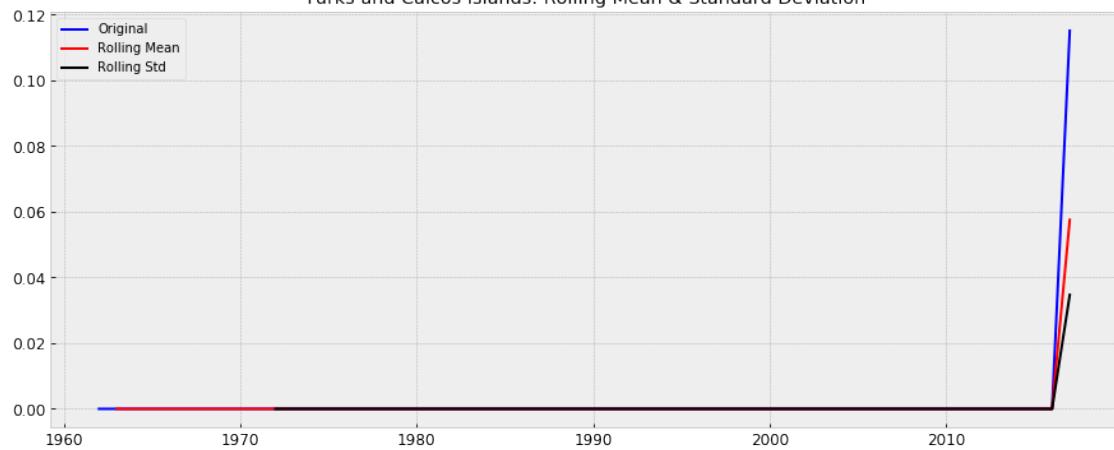
Tunisia: Rolling Mean & Standard Deviation



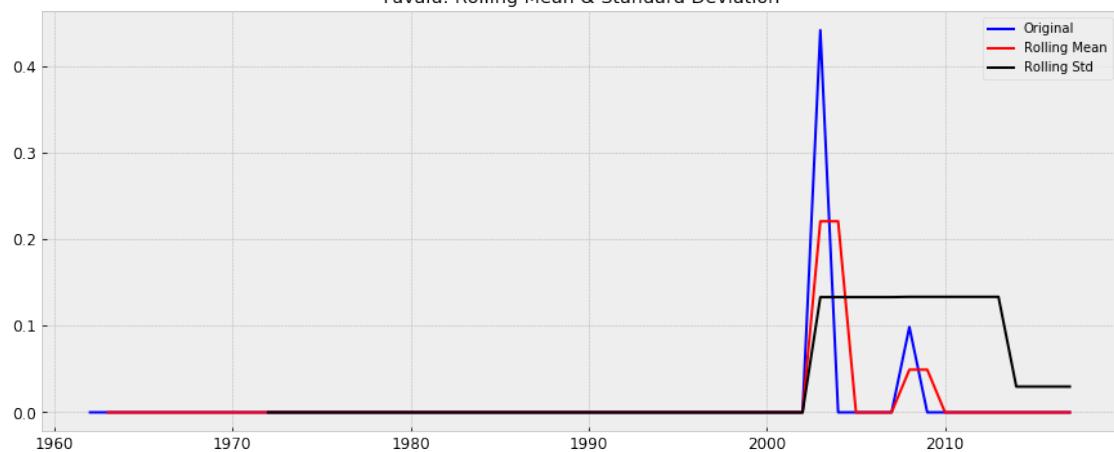
Turkey: Rolling Mean & Standard Deviation

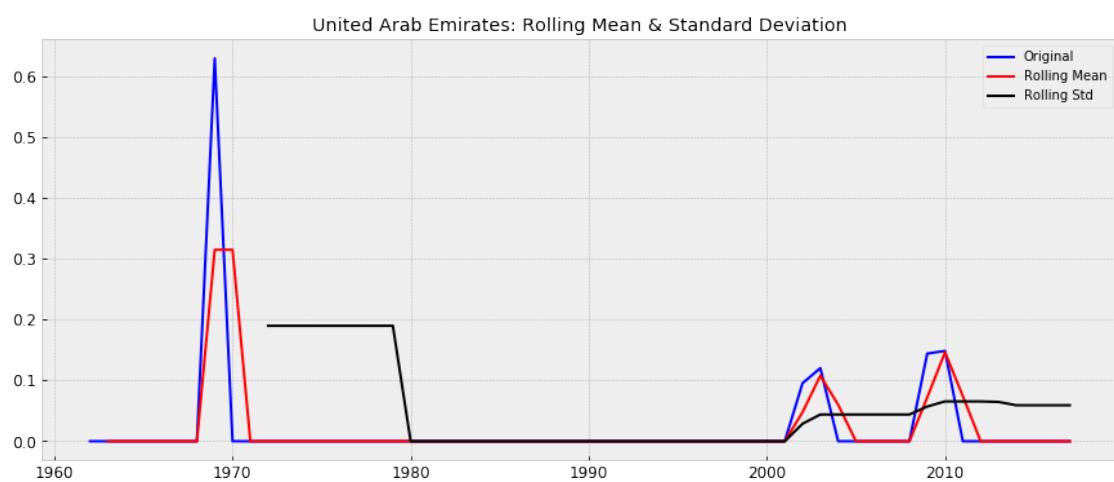
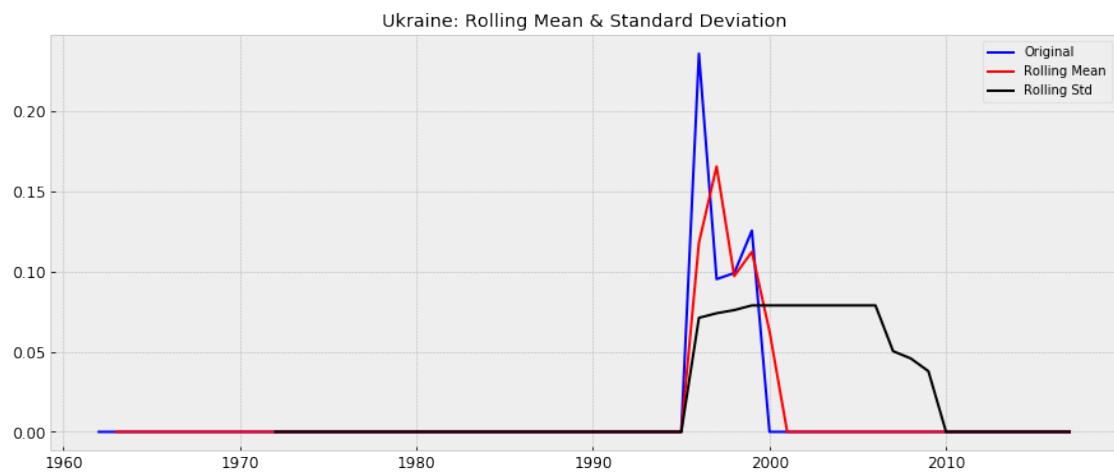


Turks and Caicos Islands: Rolling Mean & Standard Deviation

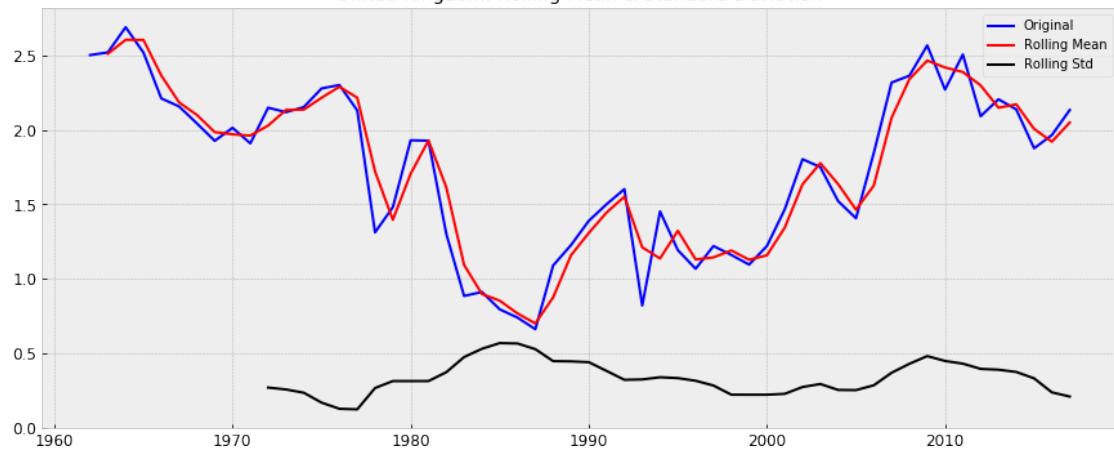


Tuvalu: Rolling Mean & Standard Deviation

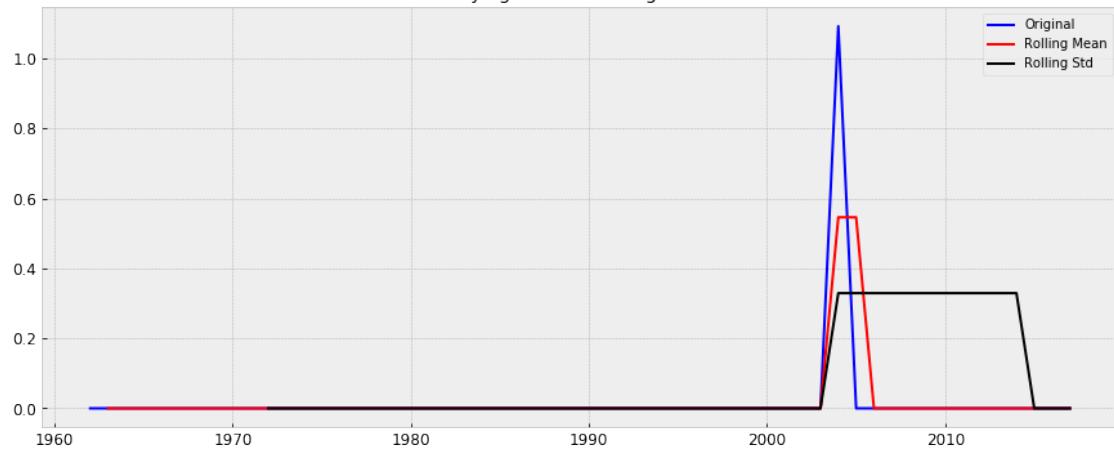




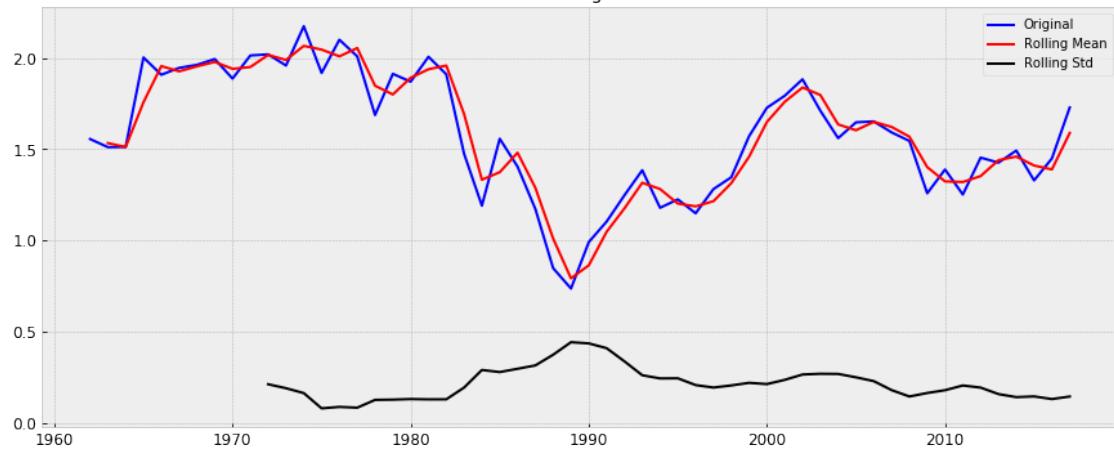
United Kingdom: Rolling Mean & Standard Deviation



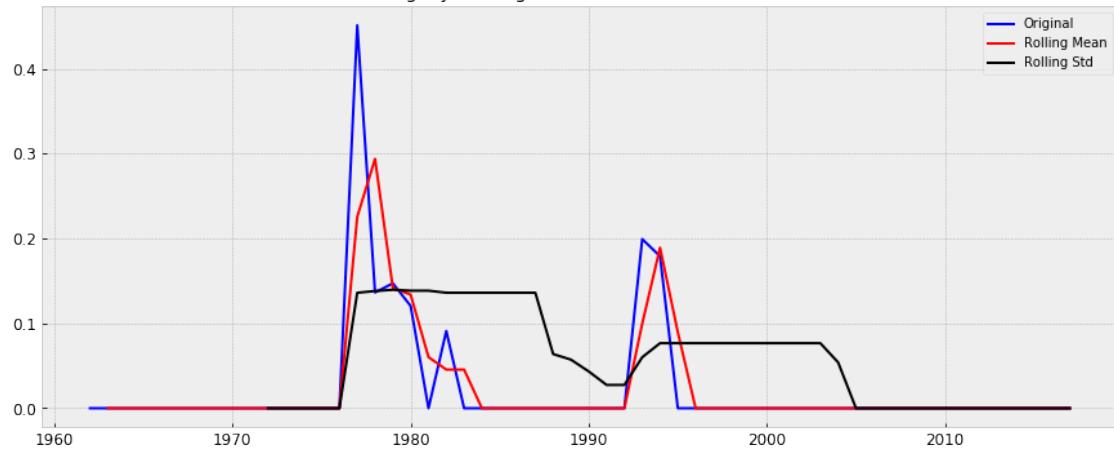
United States Minor Outlying Islands: Rolling Mean & Standard Deviation



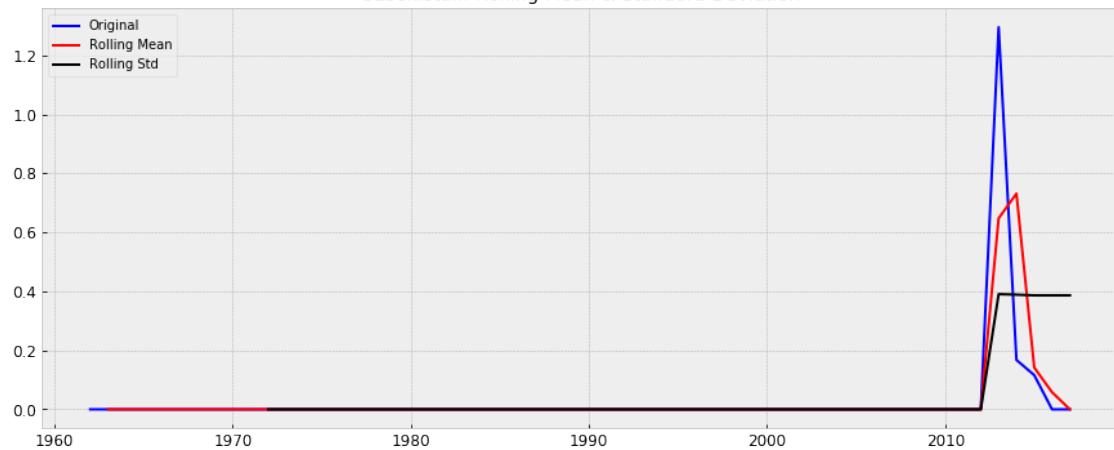
United States of America: Rolling Mean & Standard Deviation



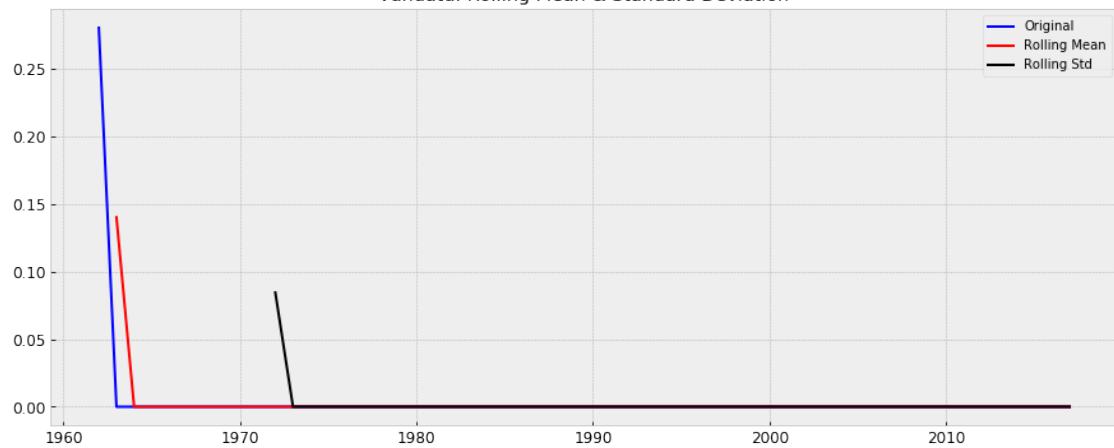
Uruguay: Rolling Mean & Standard Deviation



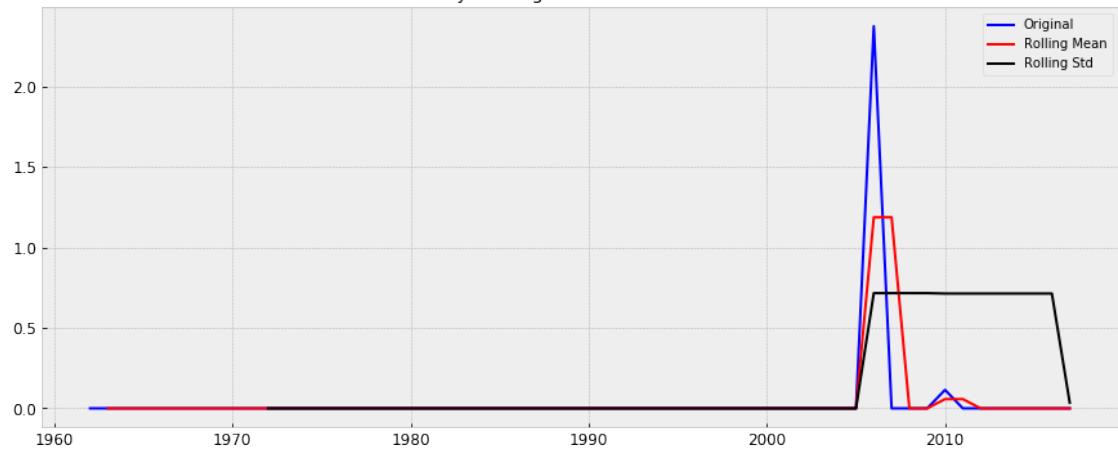
Uzbekistan: Rolling Mean & Standard Deviation



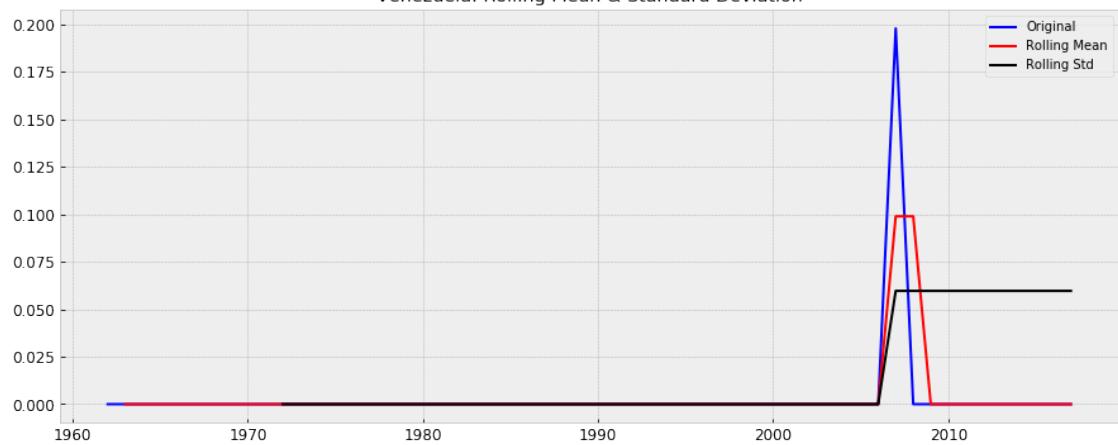
Vanuatu: Rolling Mean & Standard Deviation

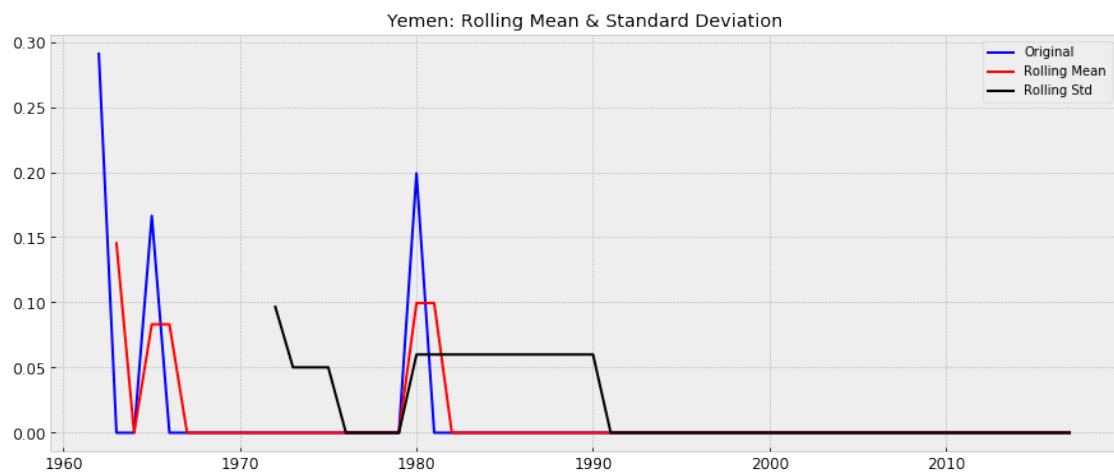
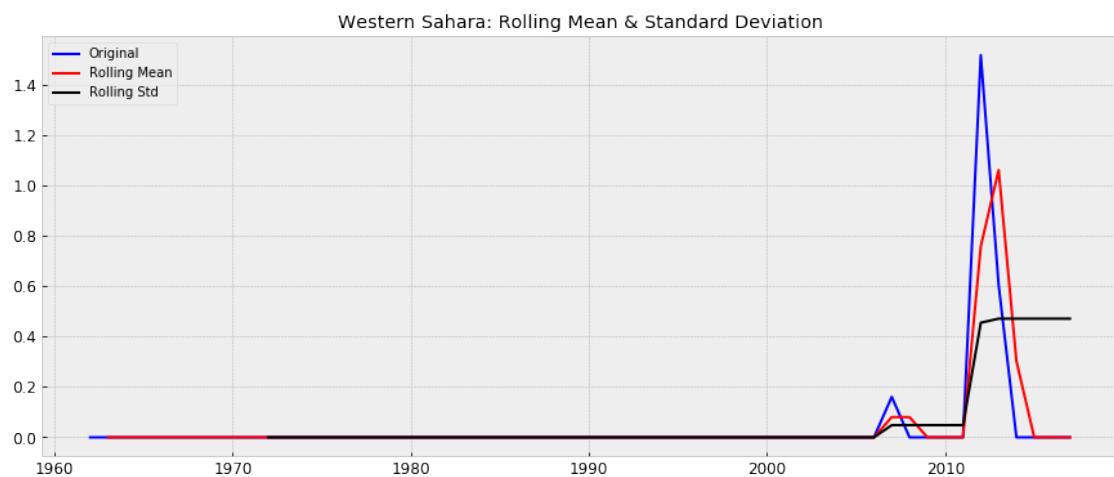
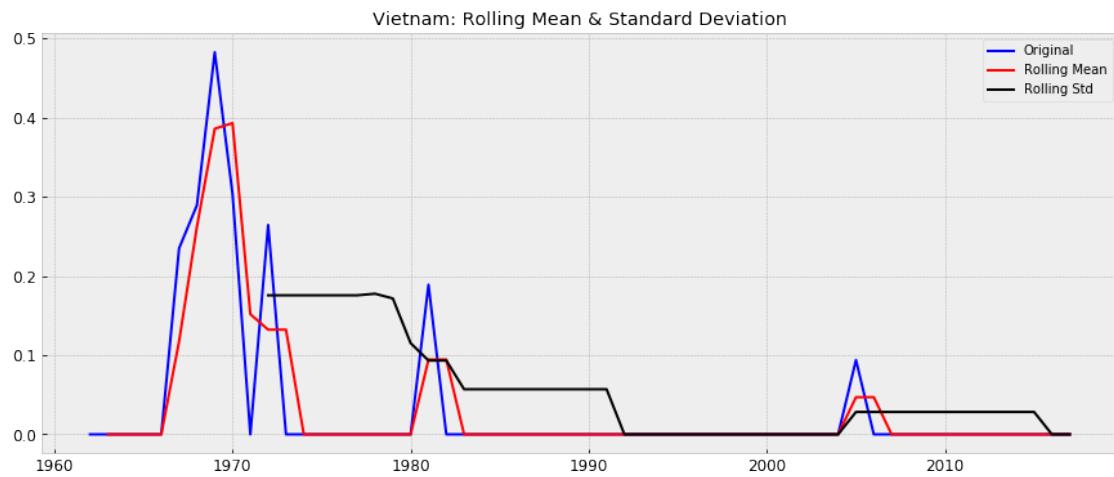


Vatican City: Rolling Mean & Standard Deviation

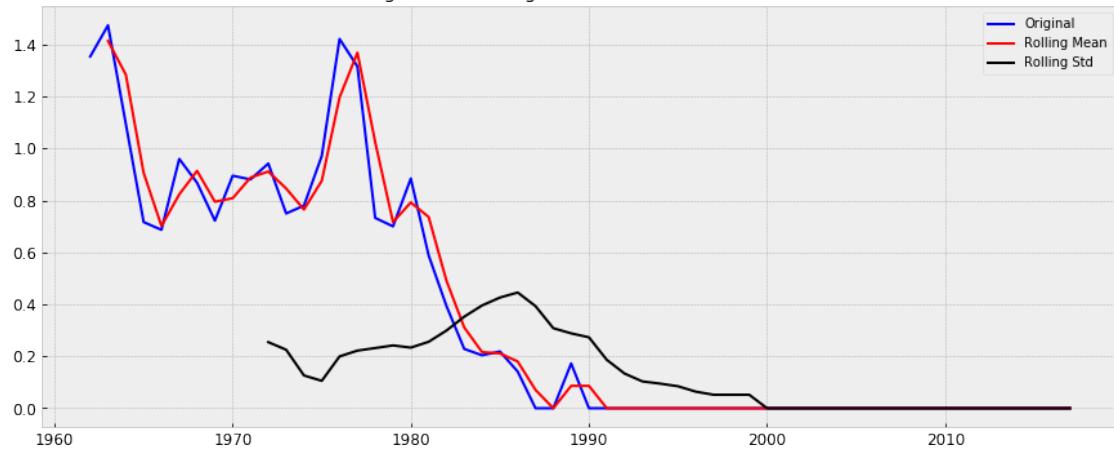


Venezuela: Rolling Mean & Standard Deviation

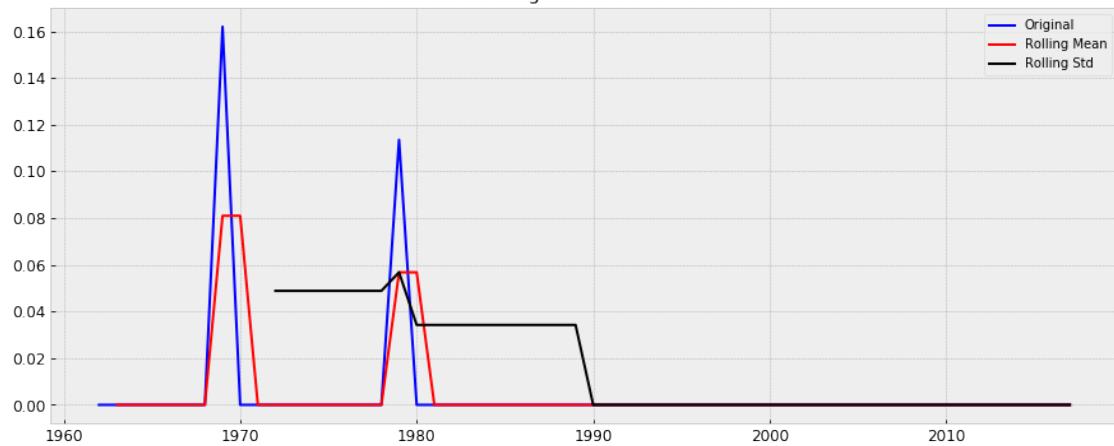




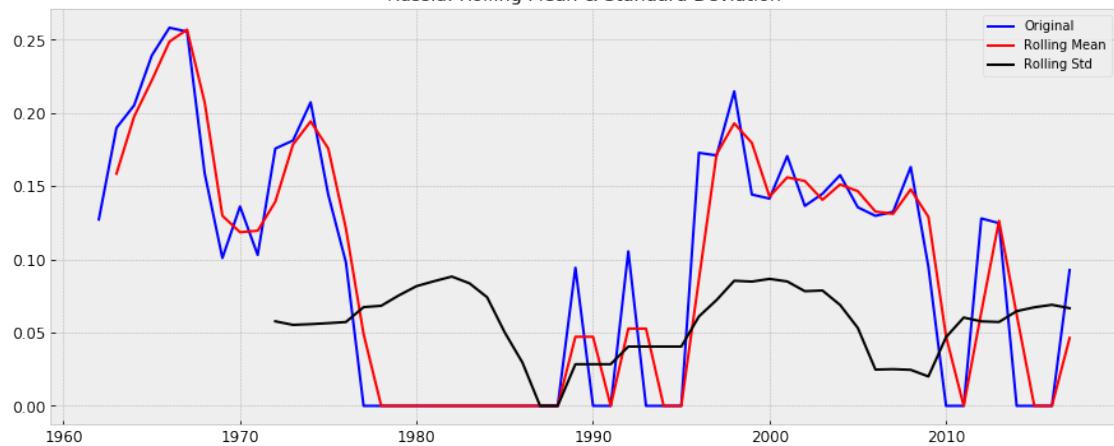
Yugoslavia: Rolling Mean & Standard Deviation



Zimbabwe: Rolling Mean & Standard Deviation



Russia: Rolling Mean & Standard Deviation



4.1.2 Augmented Dickey-Fuller (ADF) Test

- The intuition behind the test is that if the series is integrated then the lagged level of the series $y(t-1)$ will provide no relevant information in predicting the change in $y(t)$.

Null hypothesis (H₀): The time series is not stationary, it presents heterocedasticity. In other words, the time series depends on itself (i.e.: y_t depends on y_{t-1} , y_{t-1} depends on y_{t-2} - Rejecting the null hypothesis (i.e. a very low p-value) will indicate stationarity. Put simply, the t-test result is less than all critical values (1%,5%,10%)

Alternative hypothesis(H_a): the series is stationary

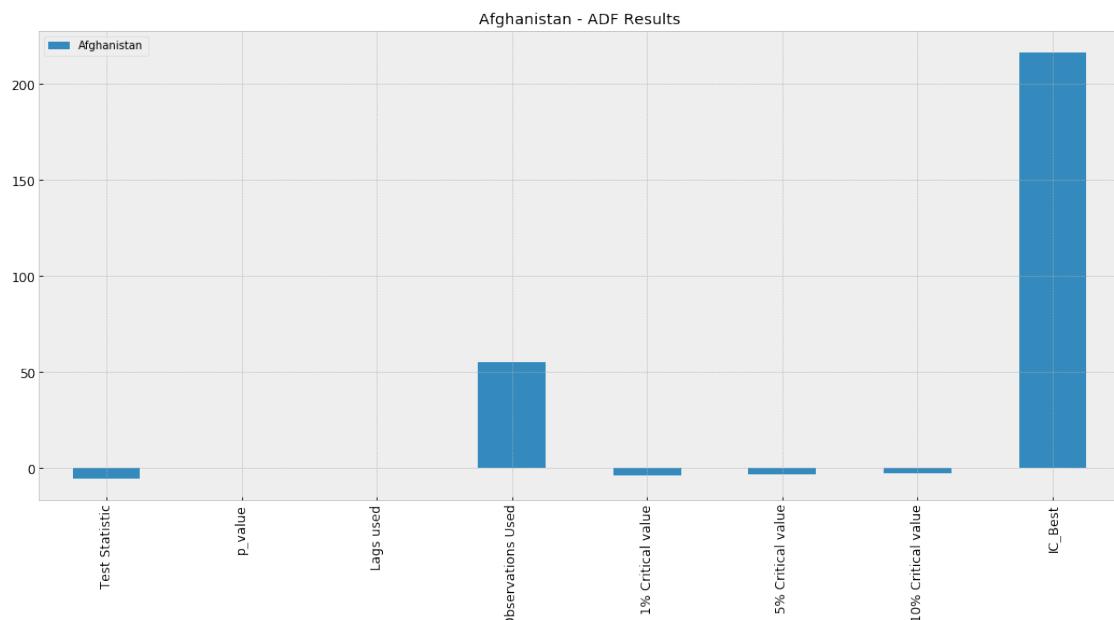
Reference ADF results description

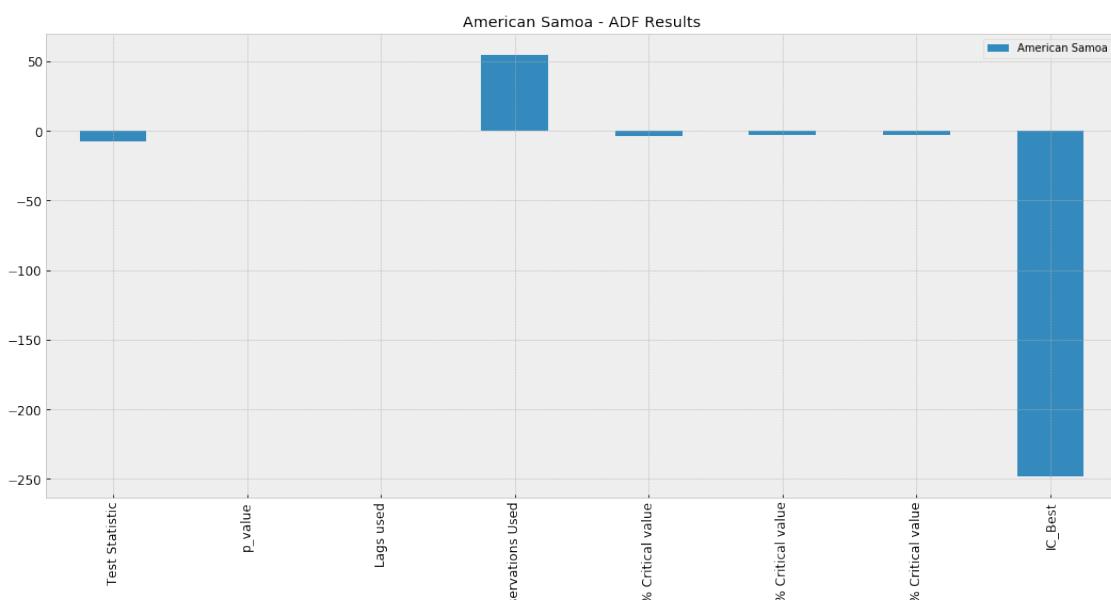
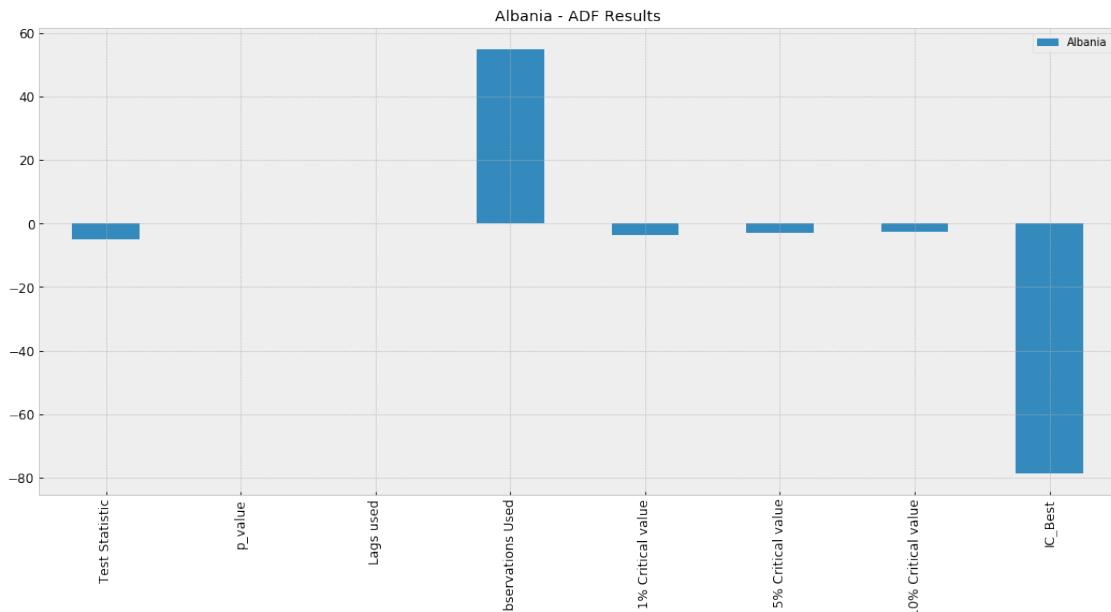
```
[179]: from statsmodels.tsa.stattools import adfuller, kpss
```

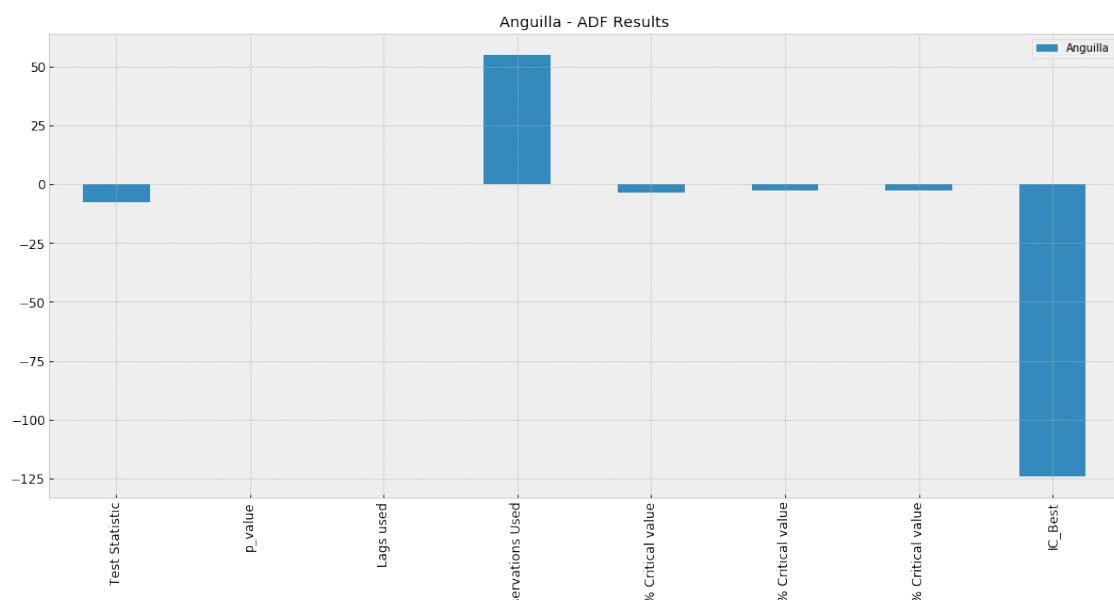
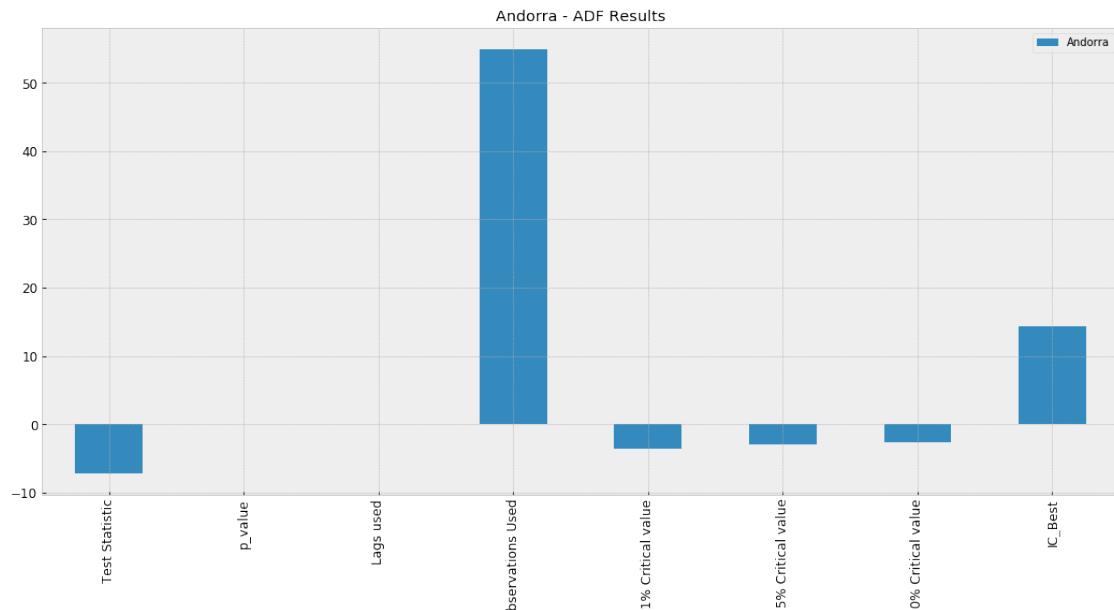
```
[180]: def ADF_Stationarity_Test(ts_groups):  
  
    #countries = list(ts_groups.columns)  
    adf_results = {}  
    col_names=['Test Statistic','p_value','Lags used','Observations Used','1%  
    ↪Critical value','5% Critical value','10% Critical value','IC_Best']  
    data = pd.DataFrame(columns=col_names,index=countries)  
  
    for country in countries:  
        #Dickey-Fuller test  
        adf_results[country] = tsa.adfuller(ts_groups[country],autolag='AIC')  
  
    for country in adf_results:  
        data.loc[country] = pd.Series({'Test Statistic':adf_results[country][0],  
                                      'p_value':adf_results[country][1],  
                                      'Lags used':adf_results[country][2],  
                                      'Observations Used':  
                                      ↪adf_results[country][3],  
                                      '1% Critical value':  
                                      ↪adf_results[country][4]['1%'],  
                                      '5% Critical value':  
                                      ↪adf_results[country][4]['5%'],  
                                      '10% Critical value':  
                                      ↪adf_results[country][4]['10%'],  
                                      'IC_Best': adf_results[country][5]  
})  
    return data.transpose()
```

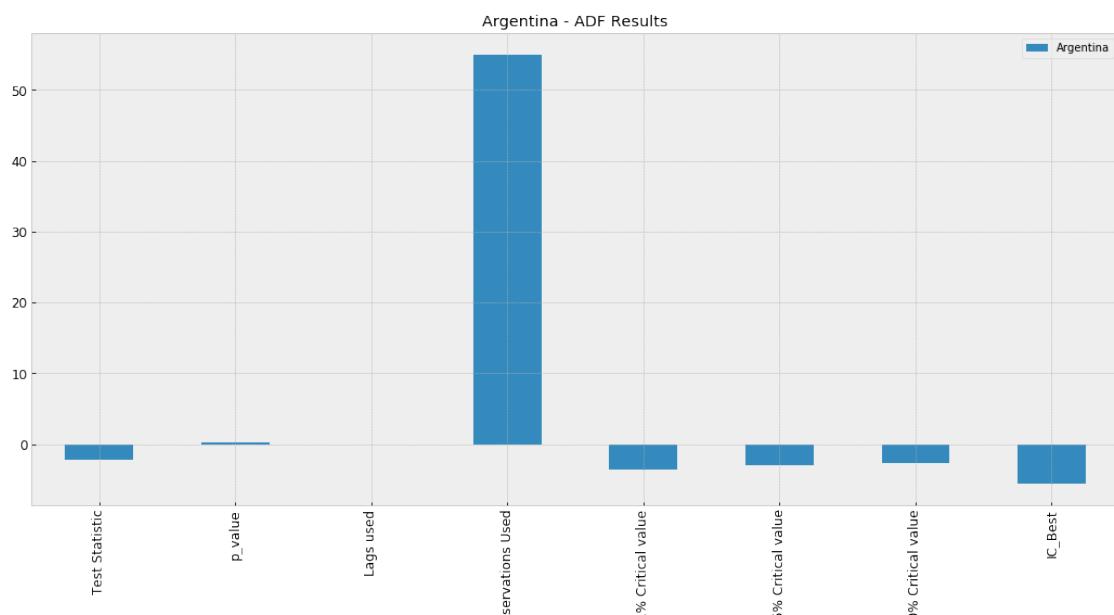
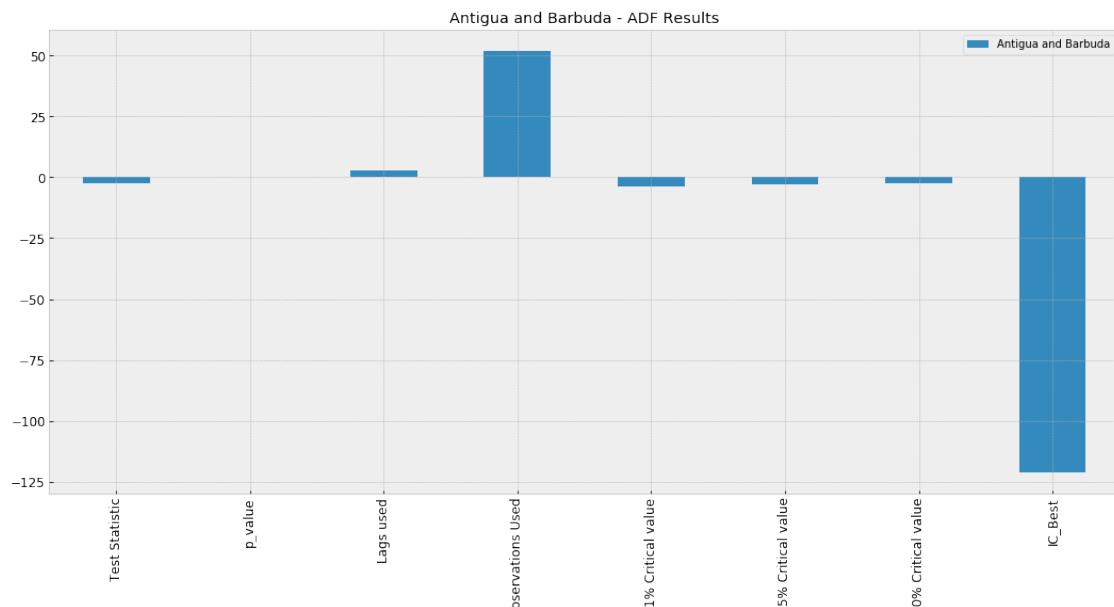
```
[181]: adf_test = ADF_Stationarity_Test(df_grps)
adf_countries = adf_test.columns
```

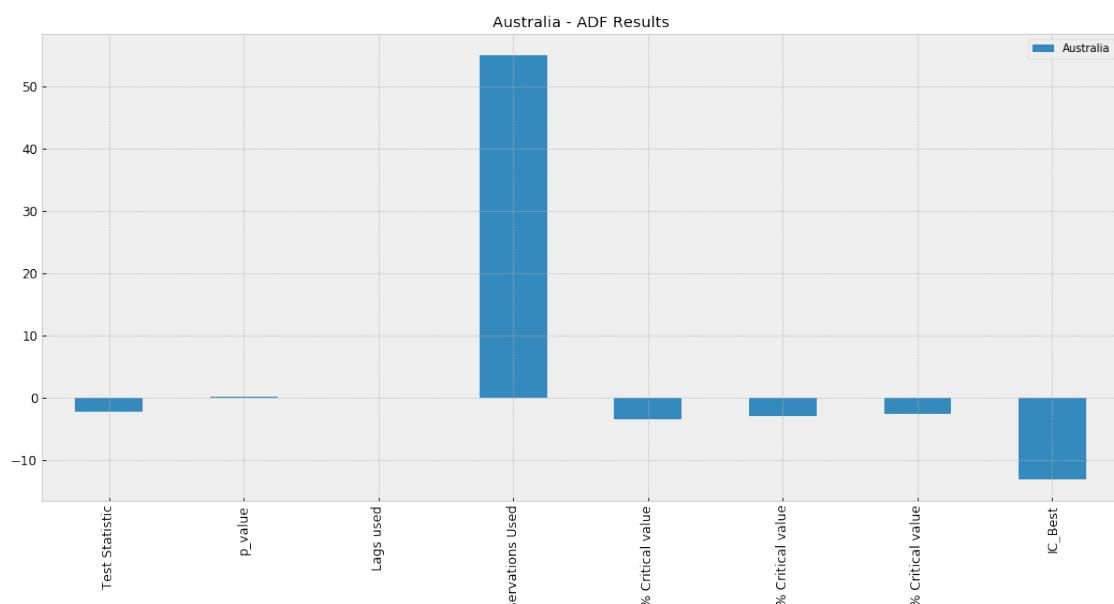
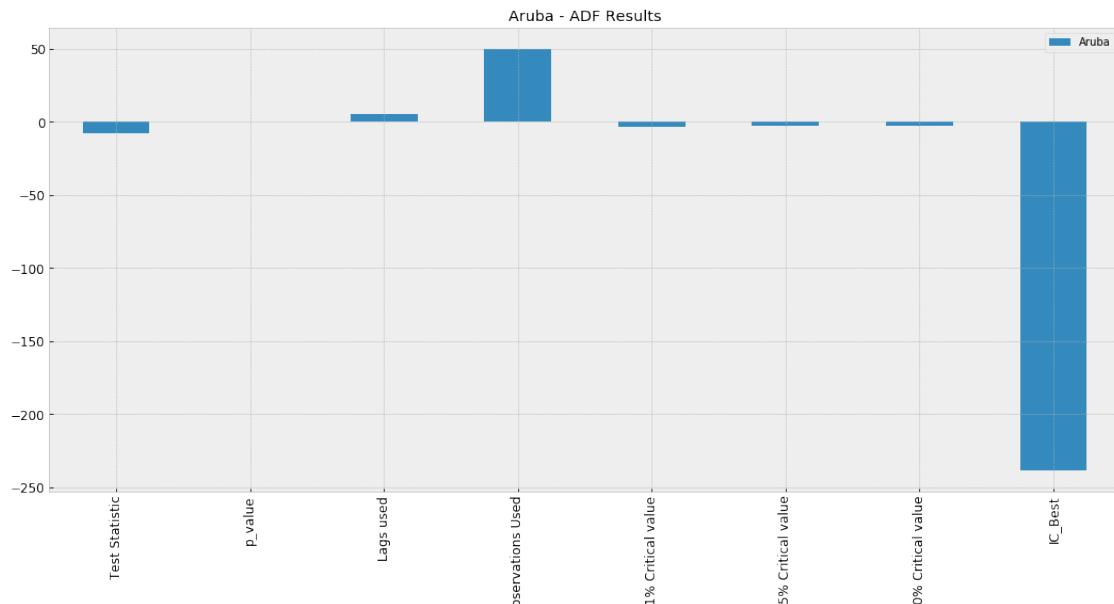
```
[182]: for country in adf_test:
    savefile = save_images(adf_test, 'adf_test')
    adf_test[country].plot.bar()
    plt.legend(loc='best')
    plt.title('{} - ADF Results'.format(adf_test[country].name))
    plt.savefig(str('{}_ADF.png'.format(savefile)))
    plt.show()
```

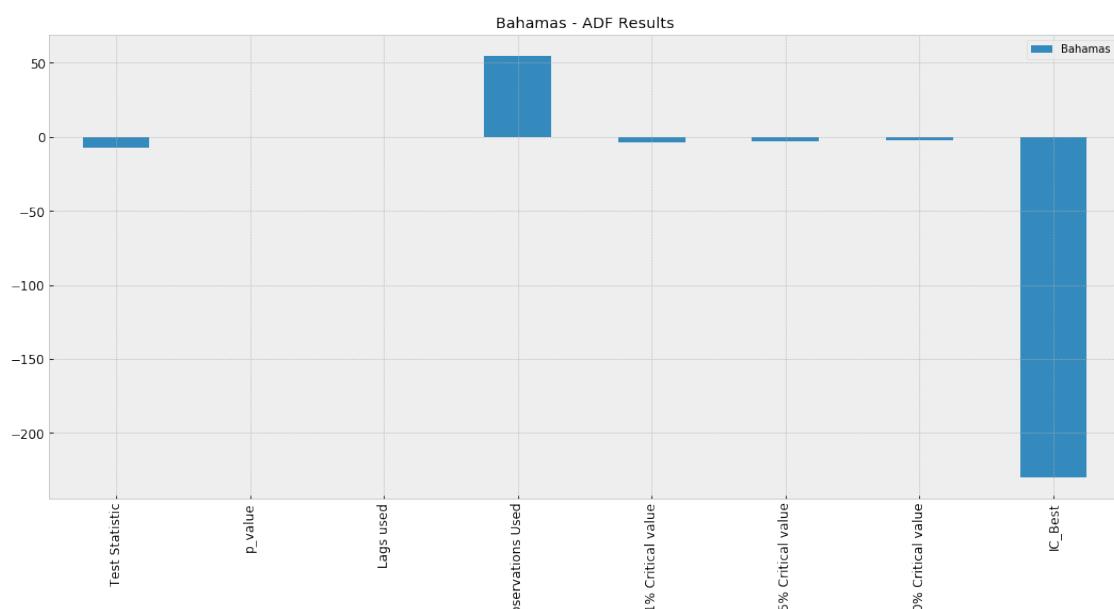
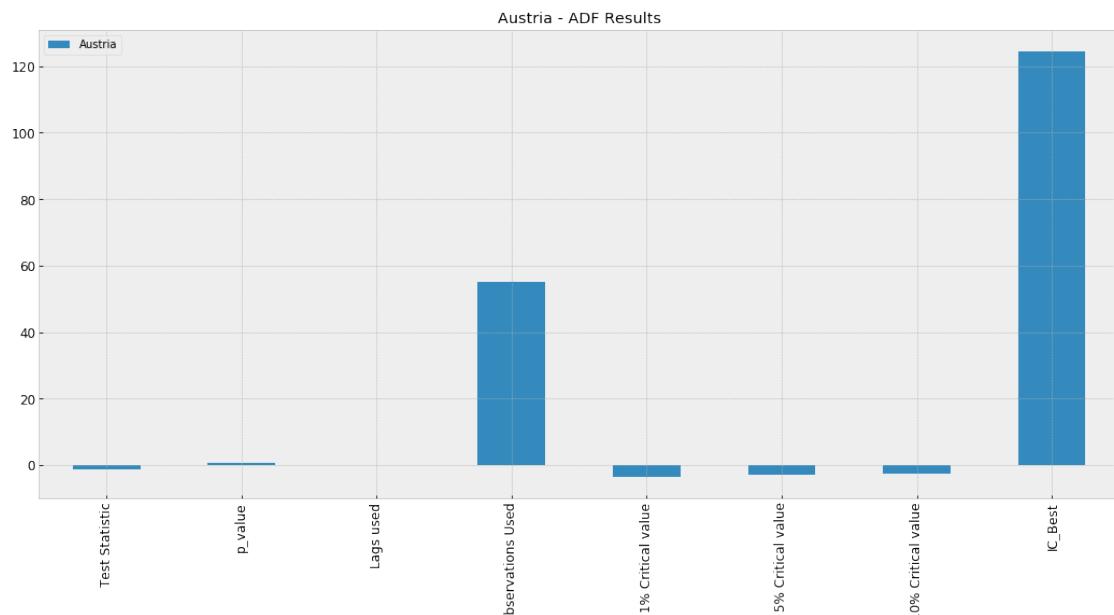


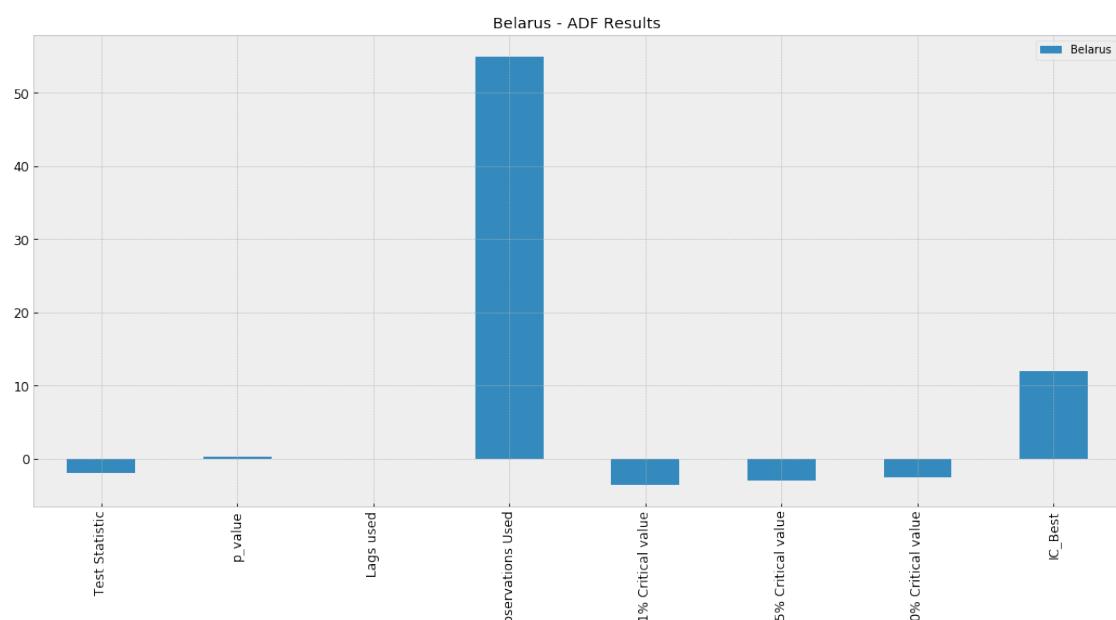
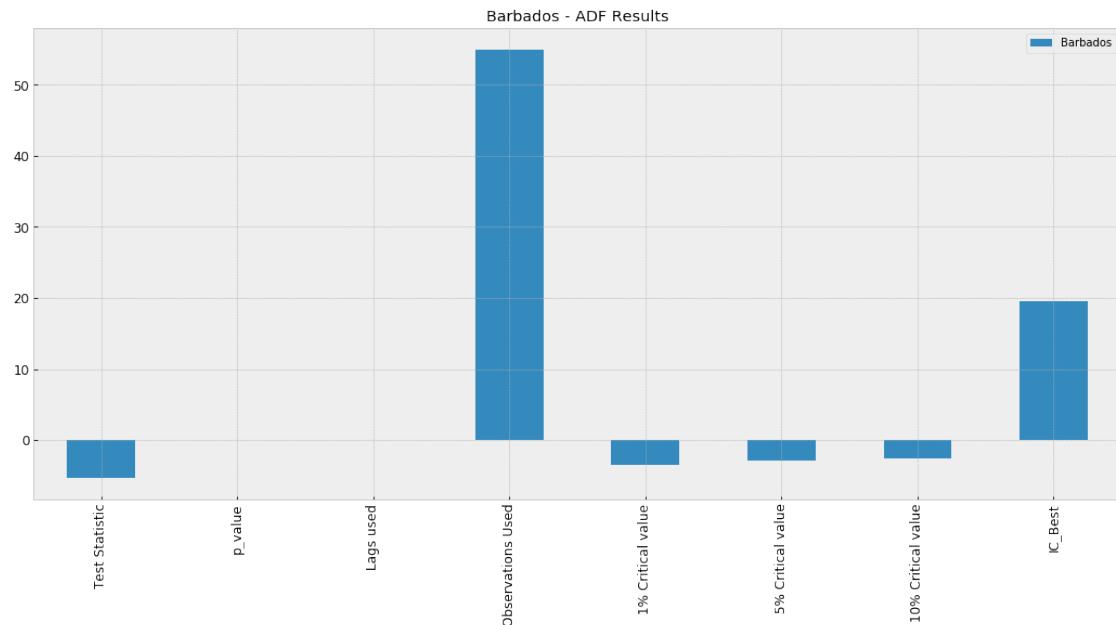


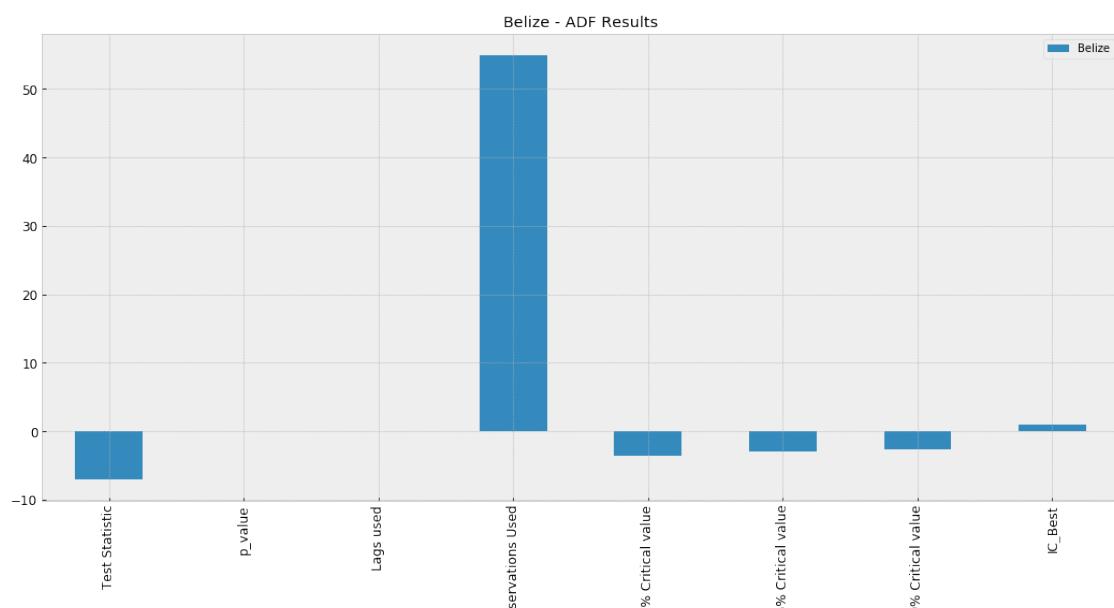
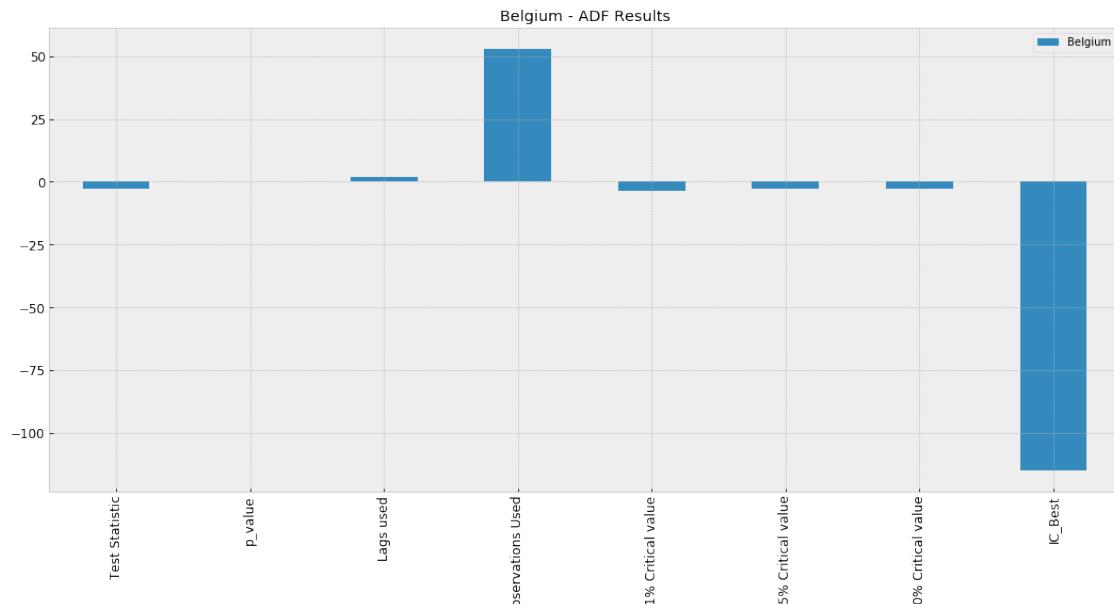


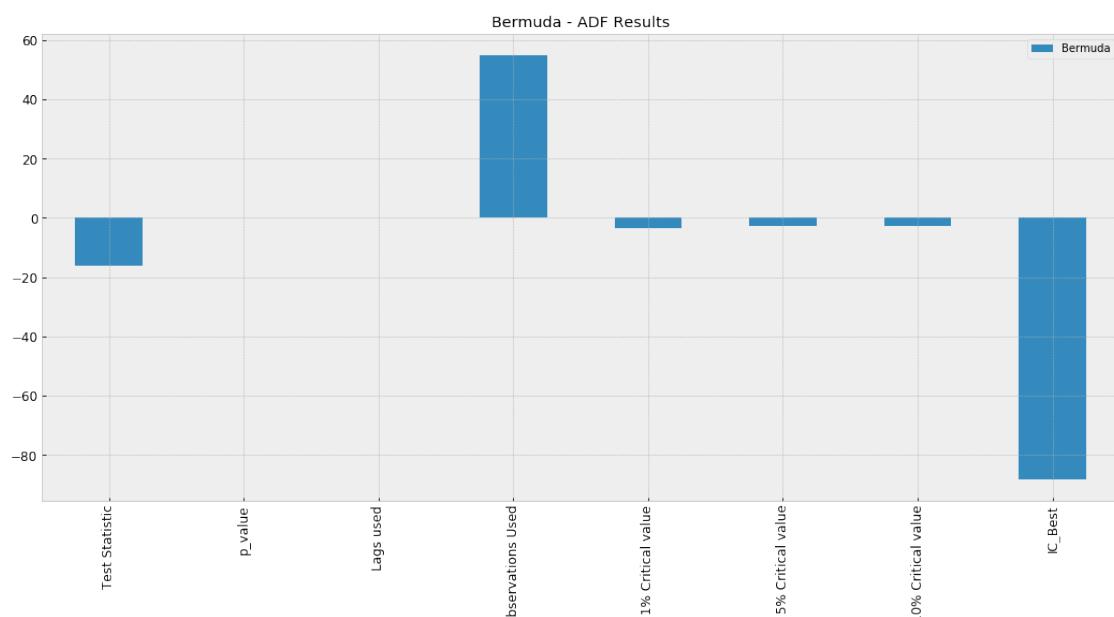
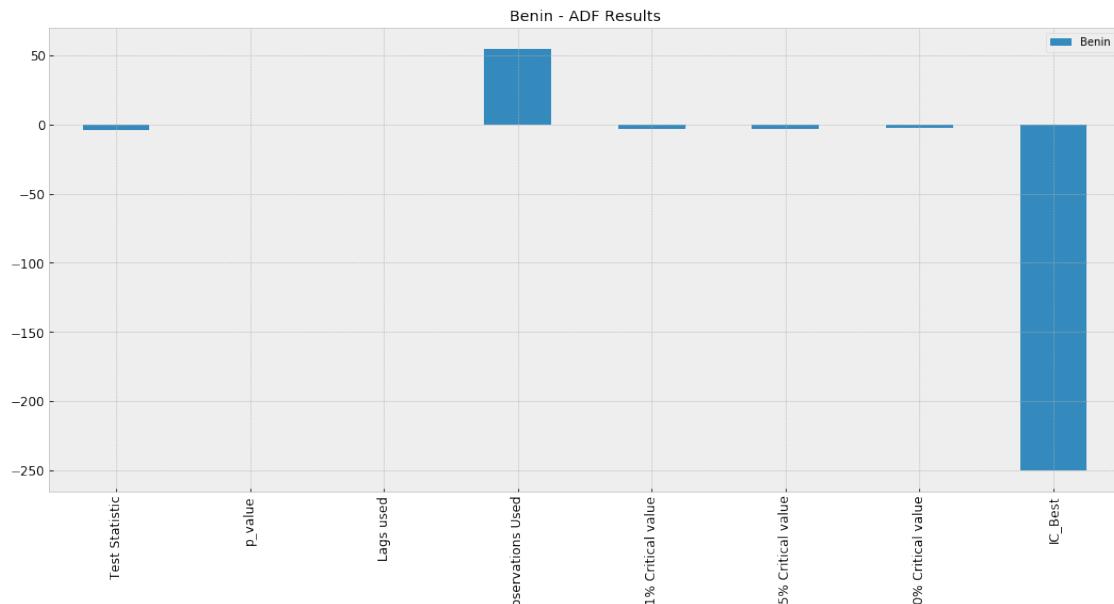


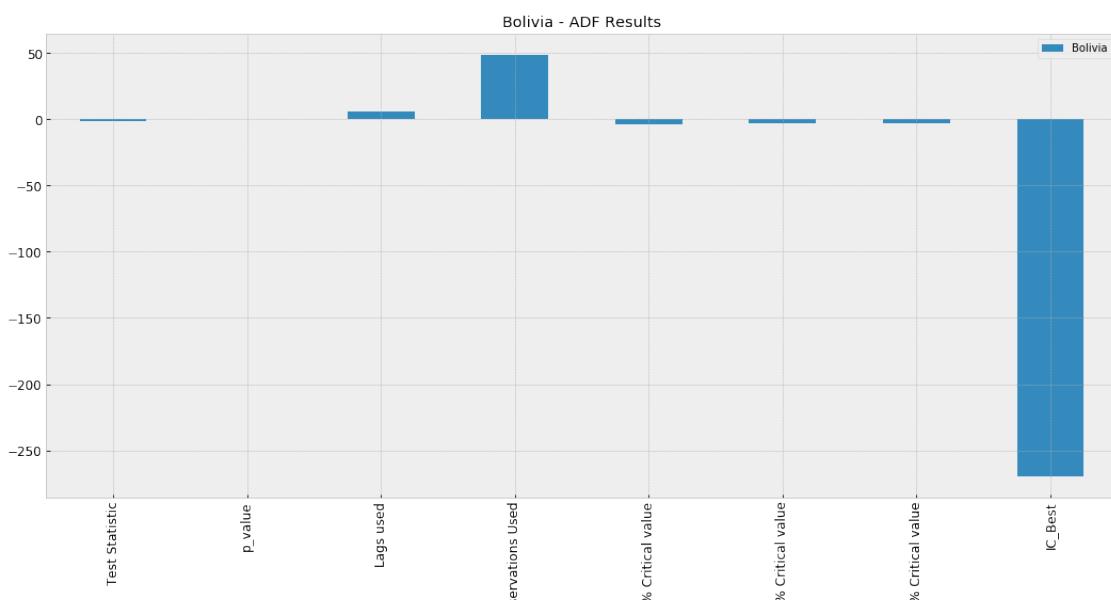
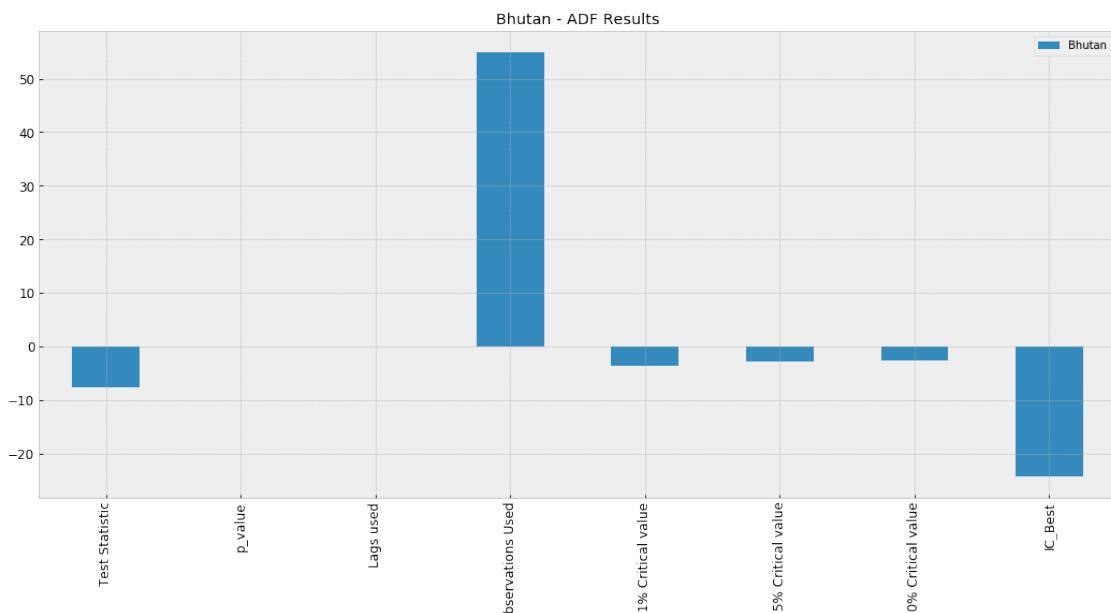


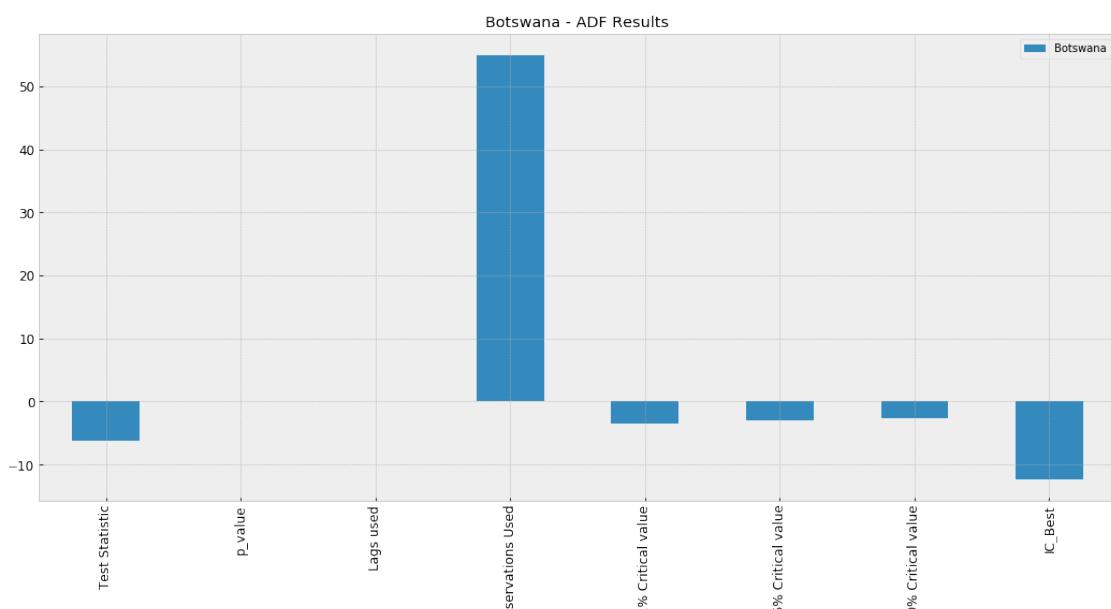
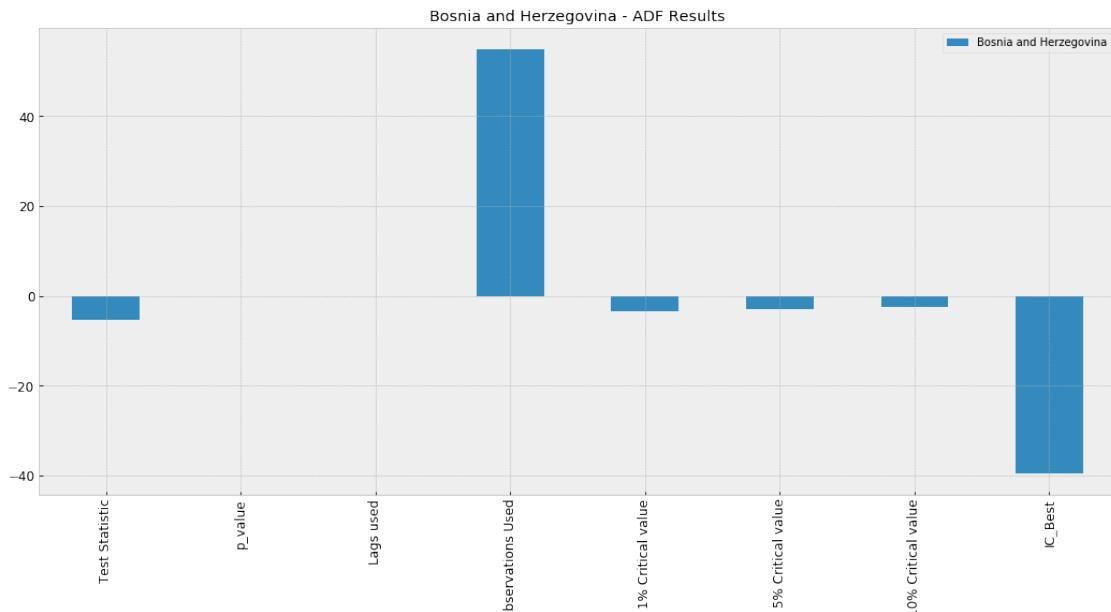


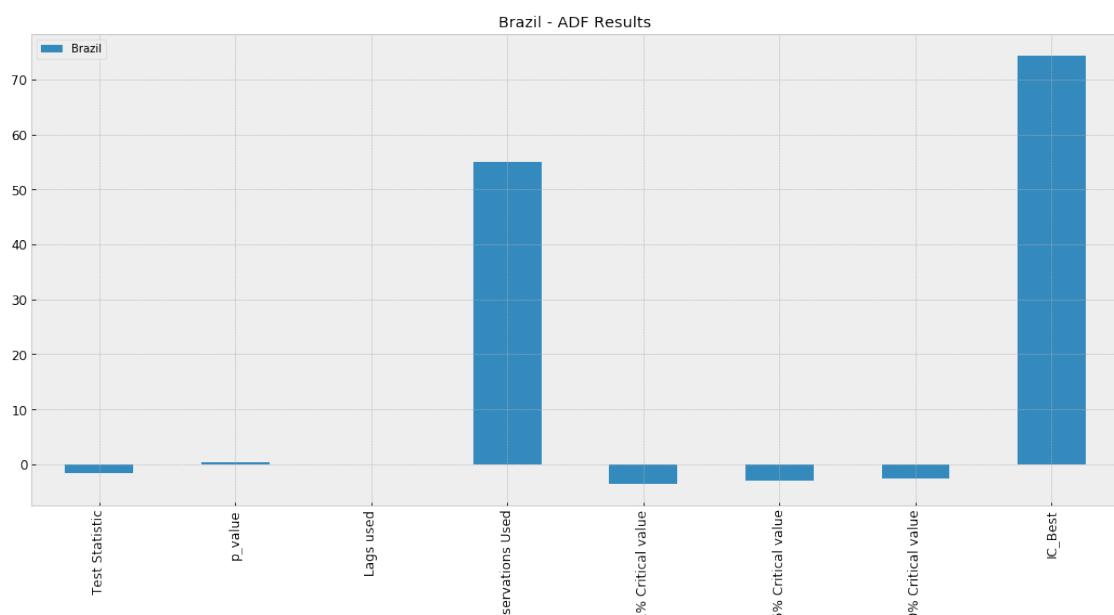
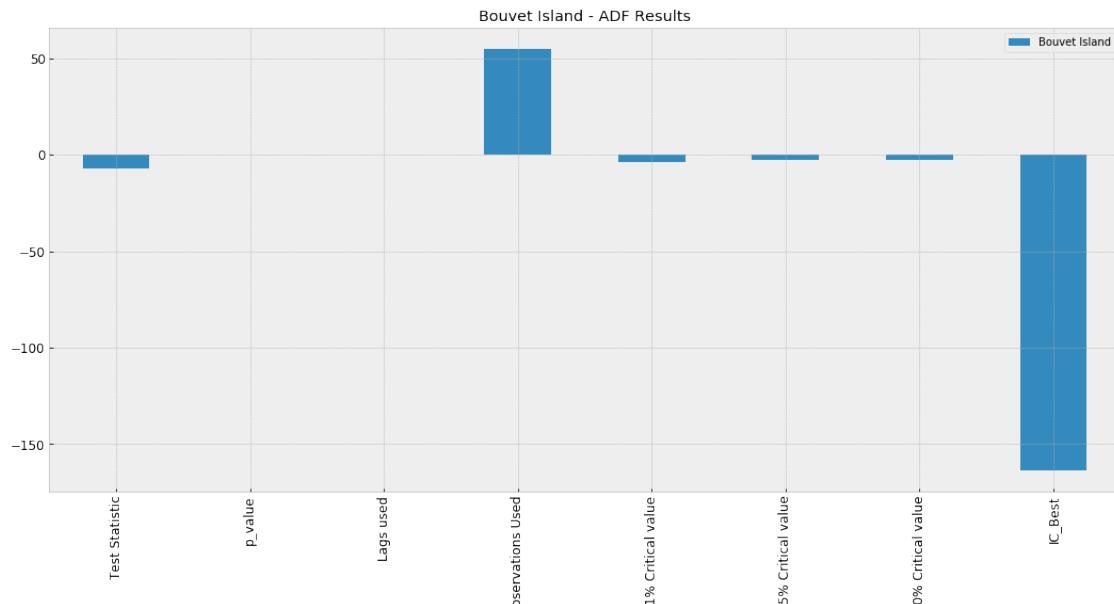


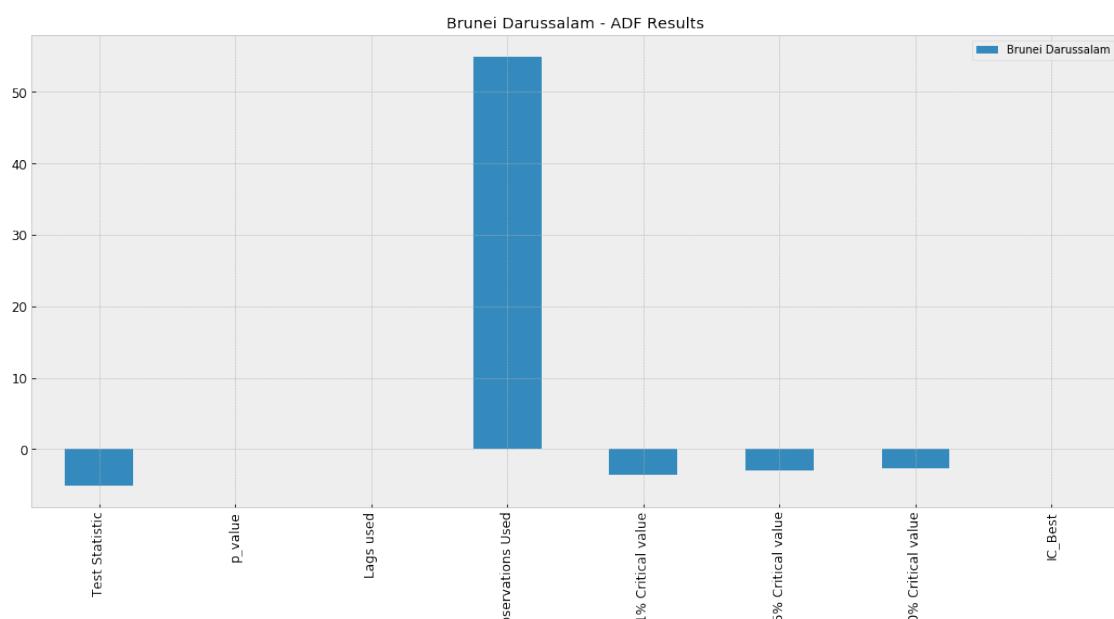
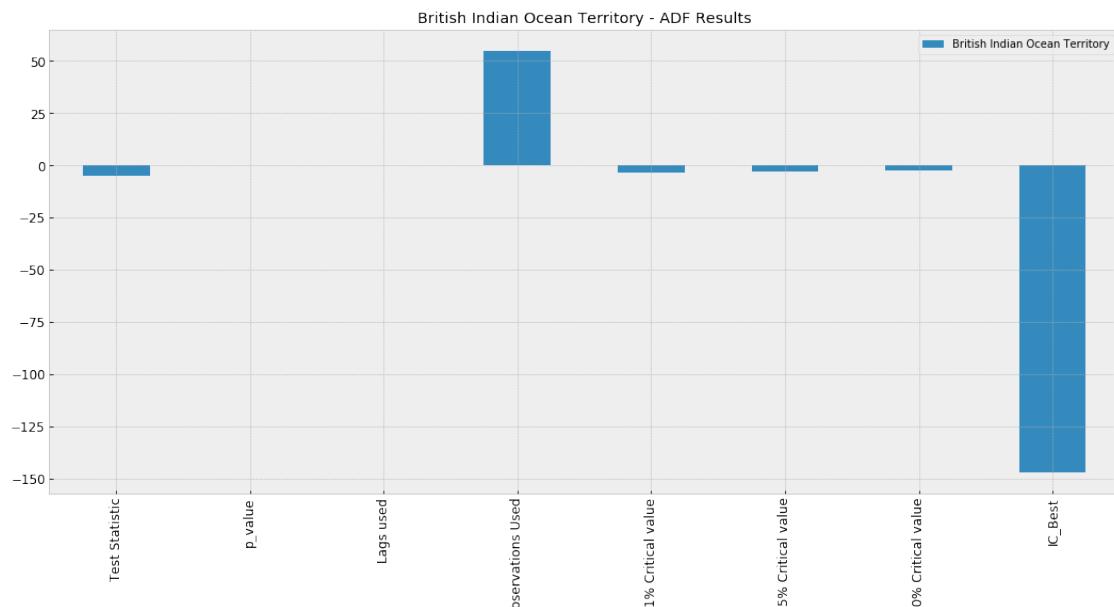


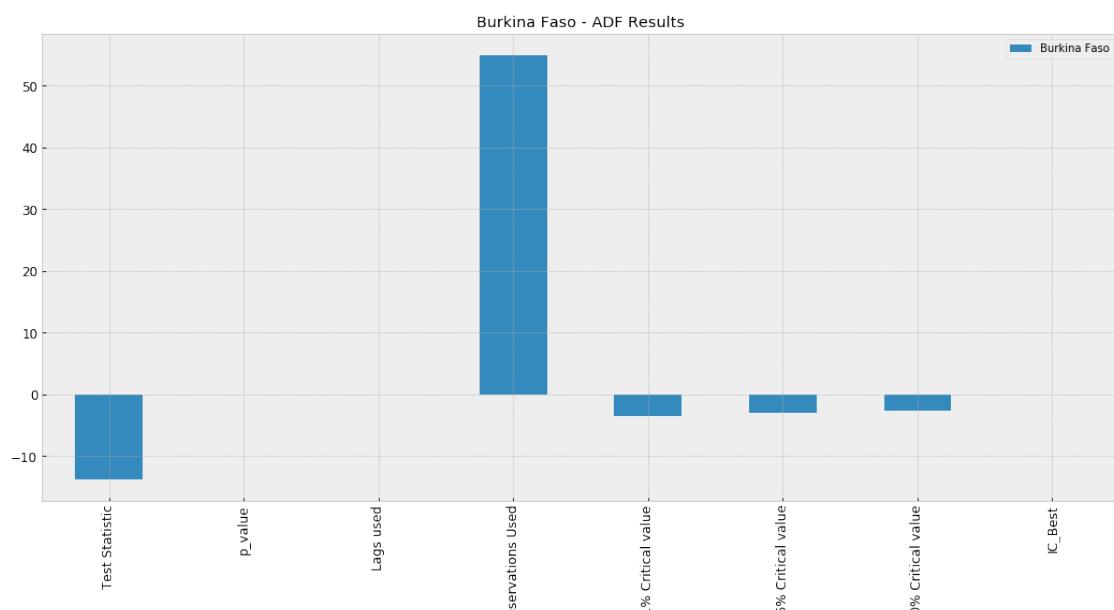
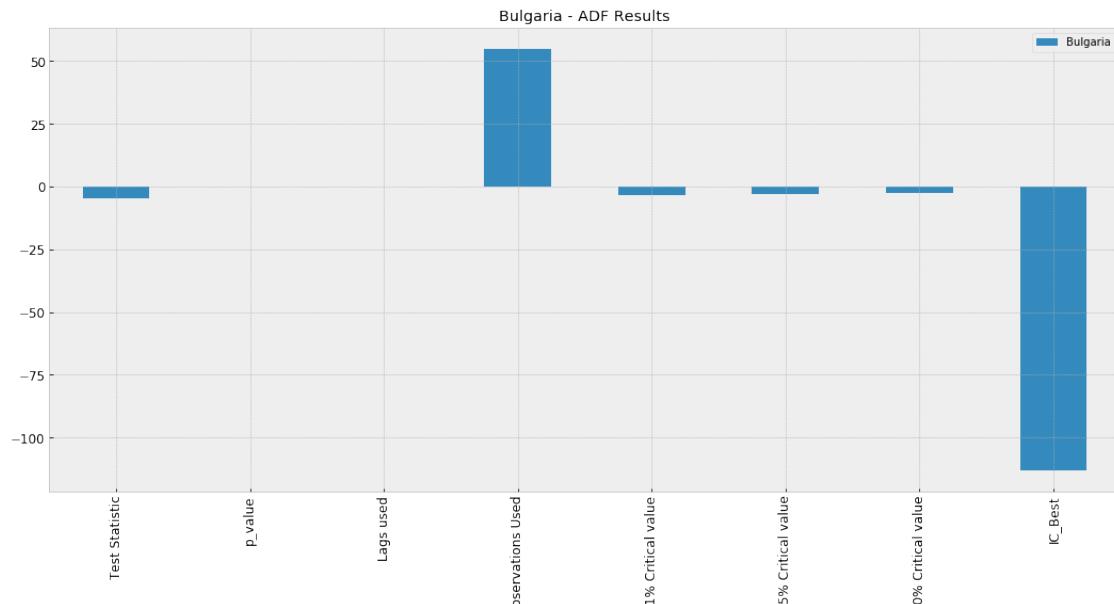


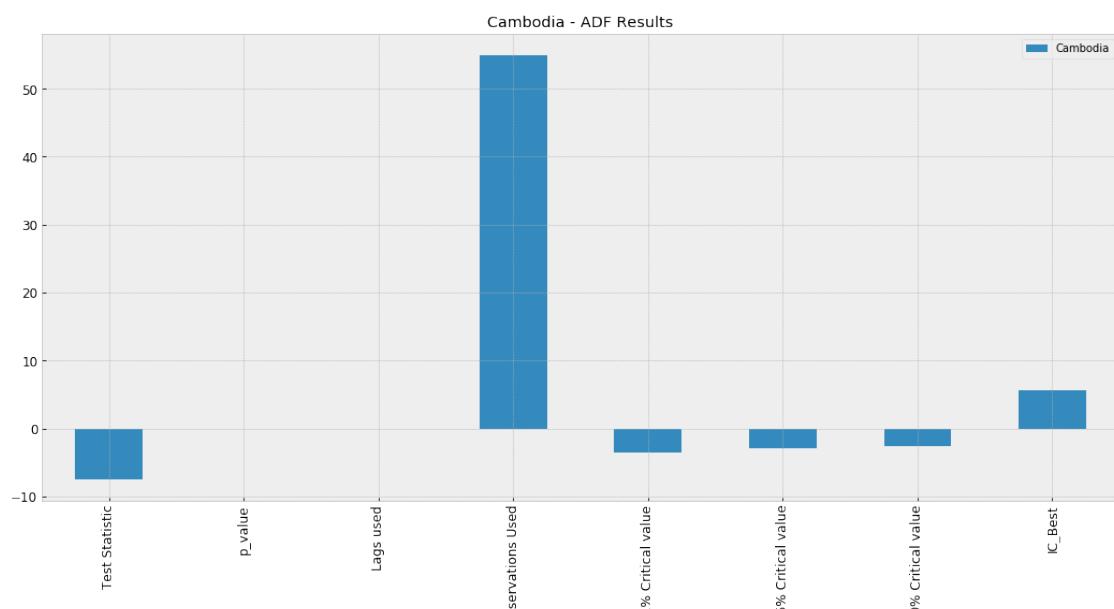
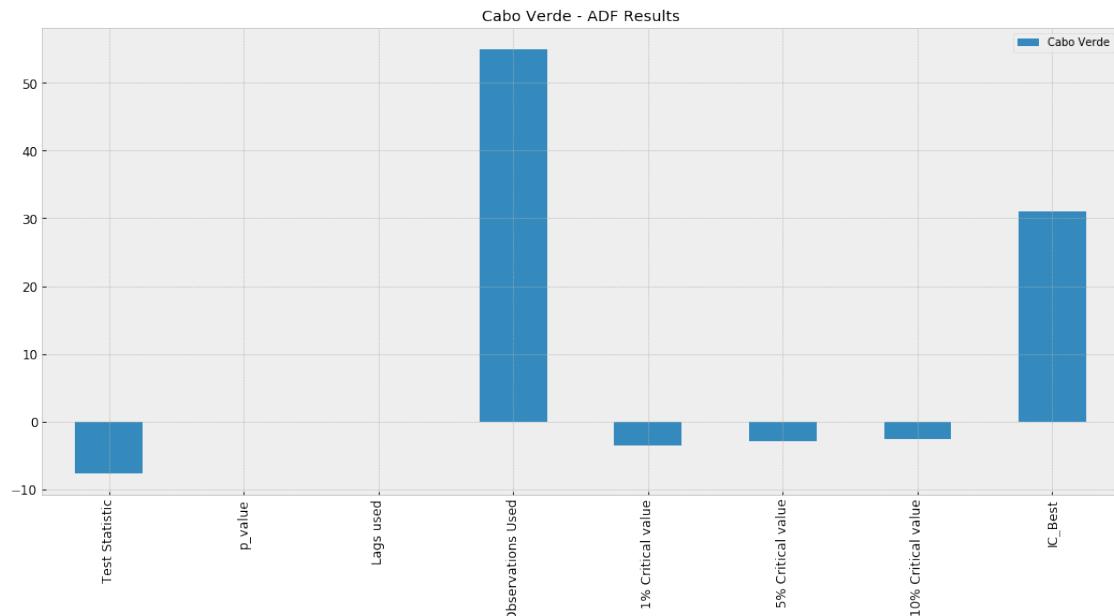


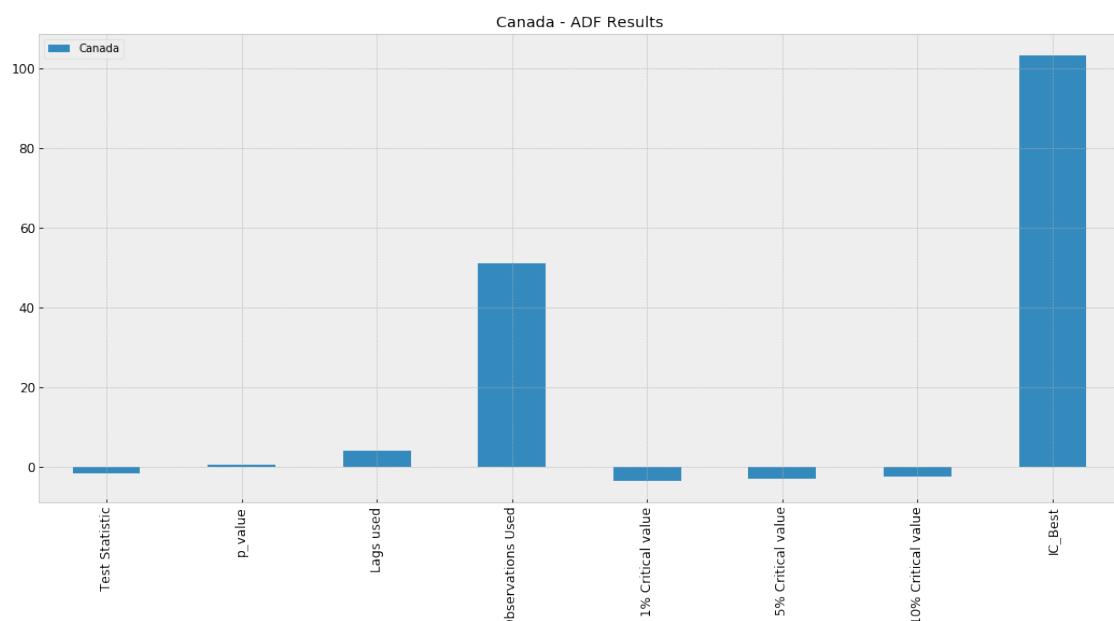
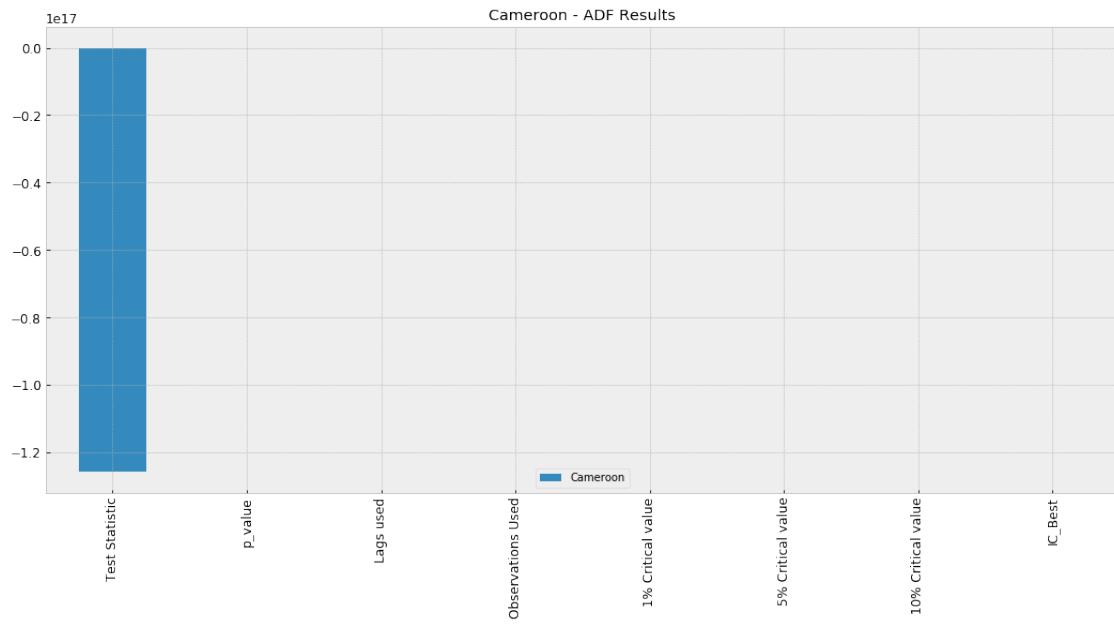


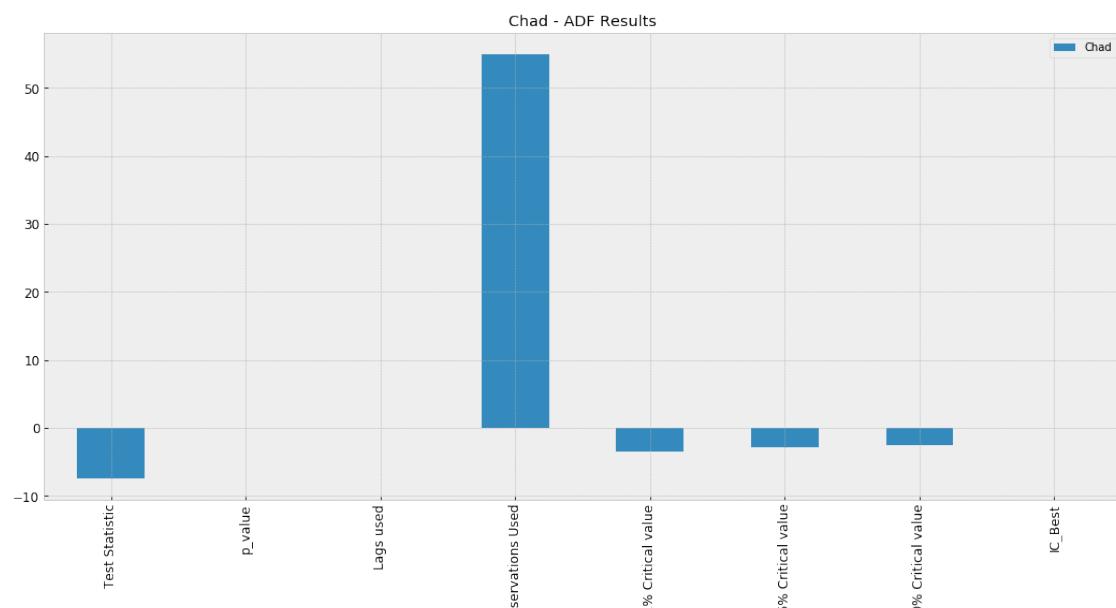
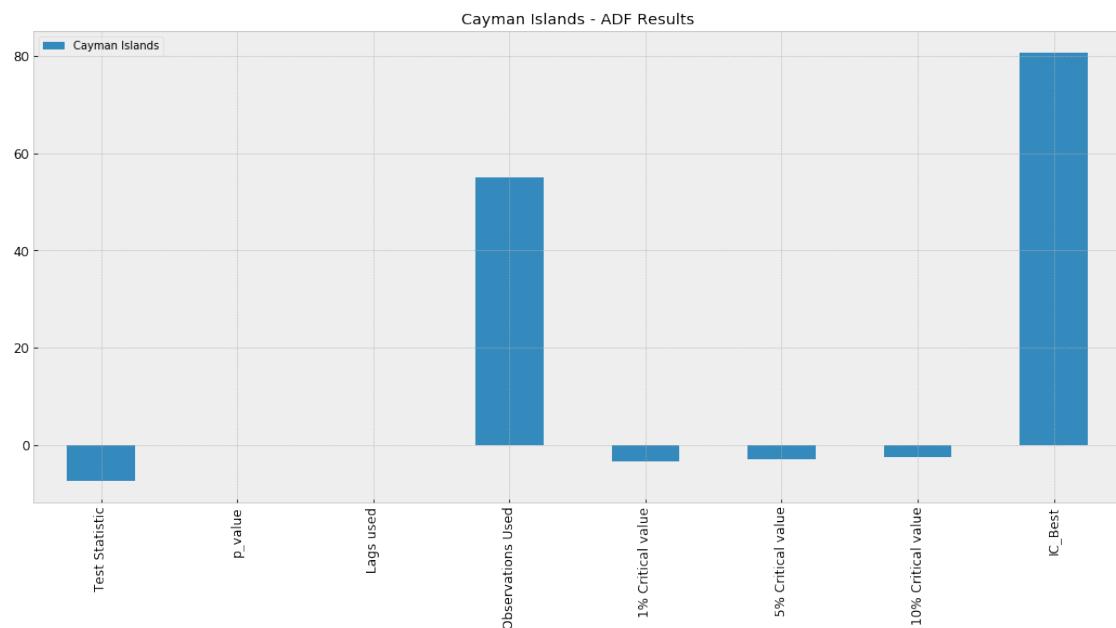


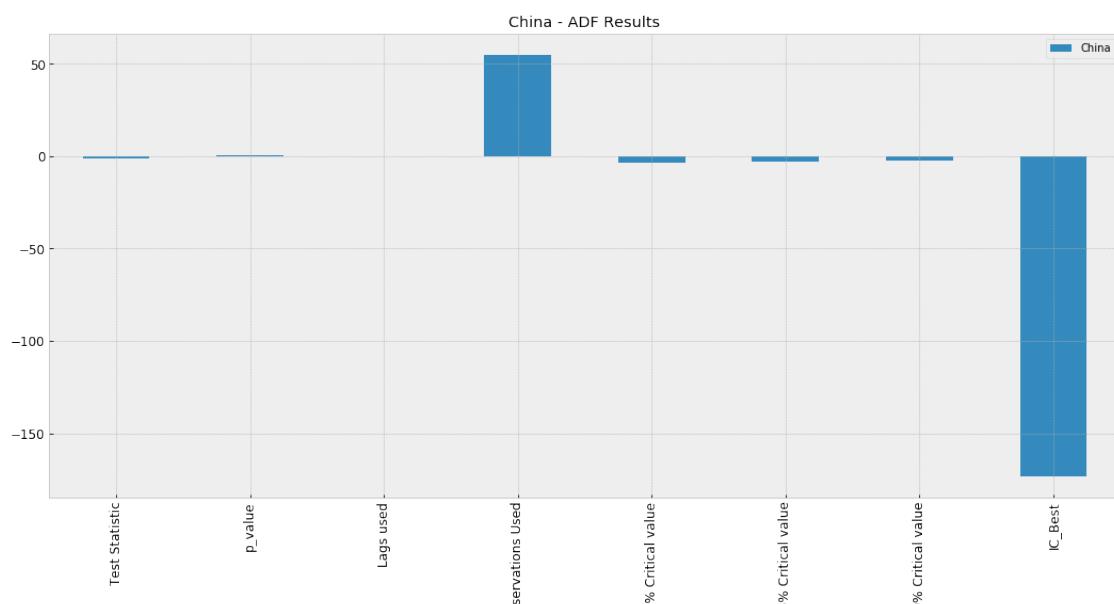
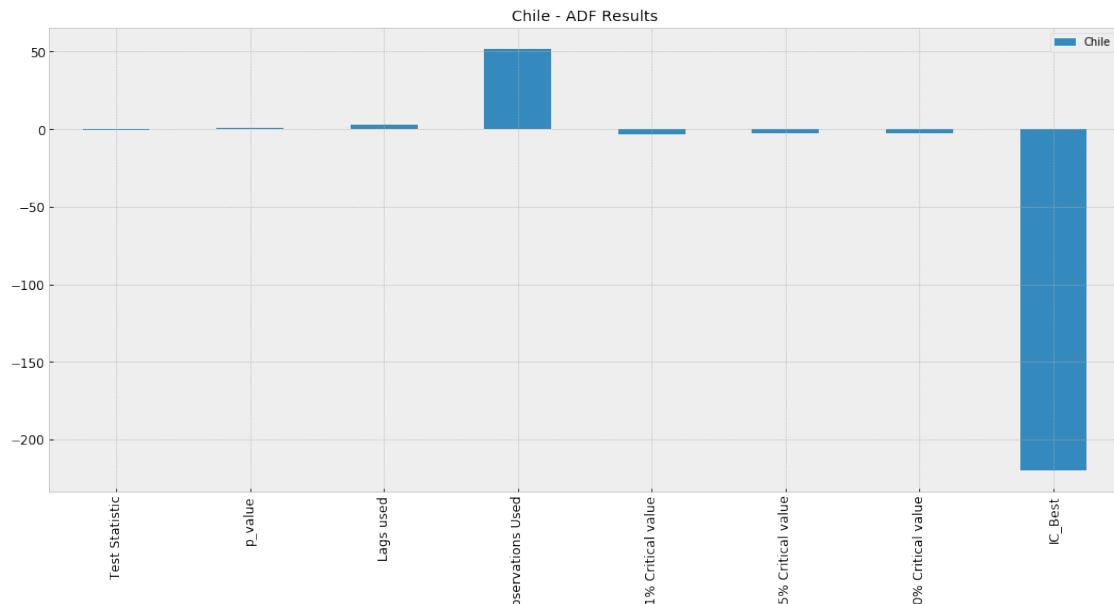


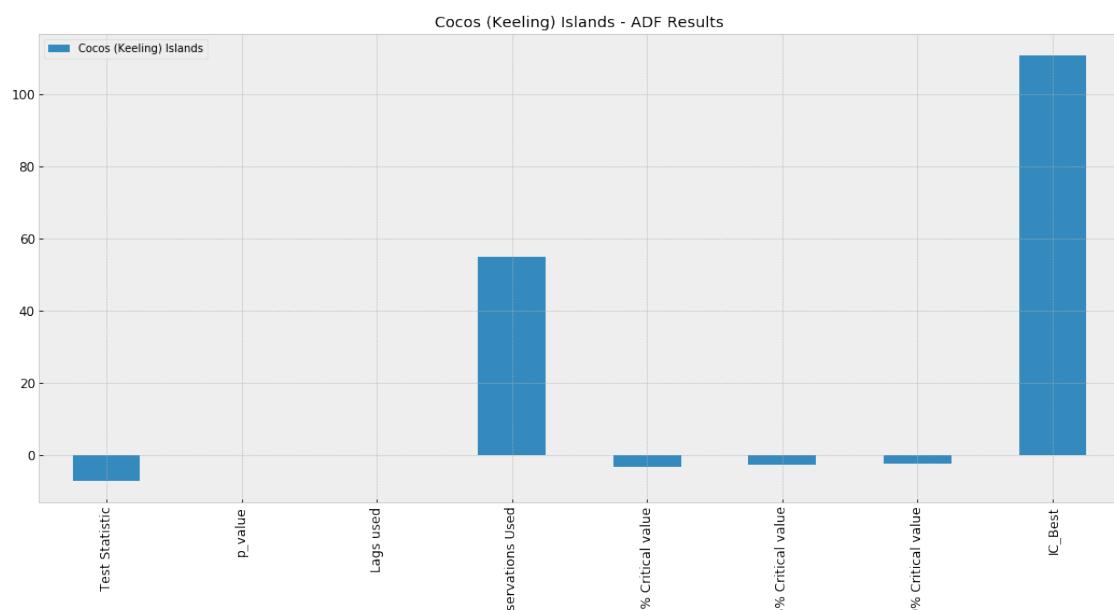
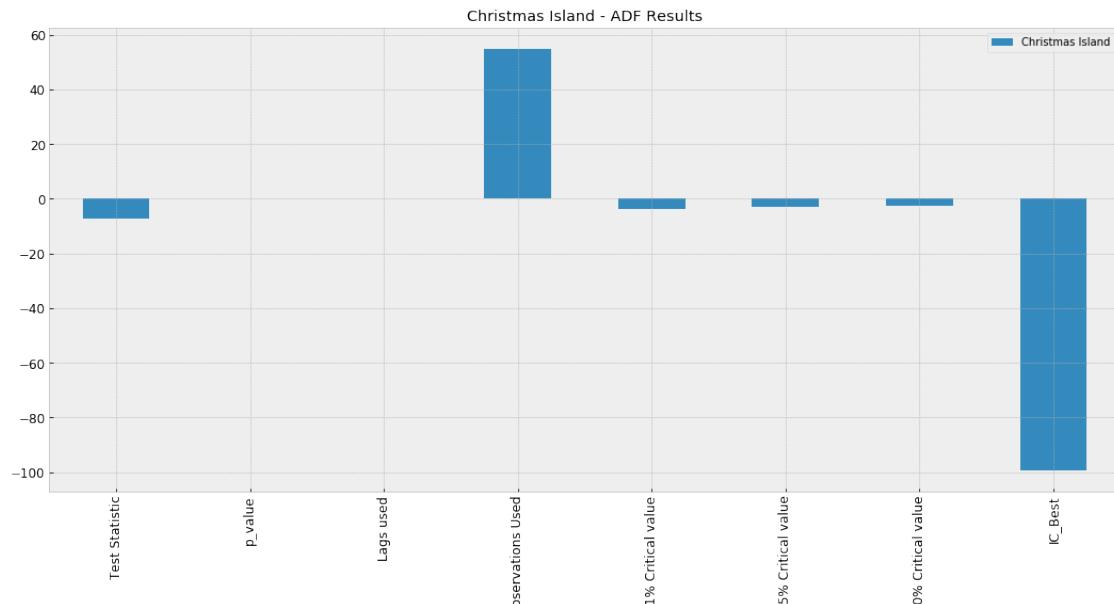


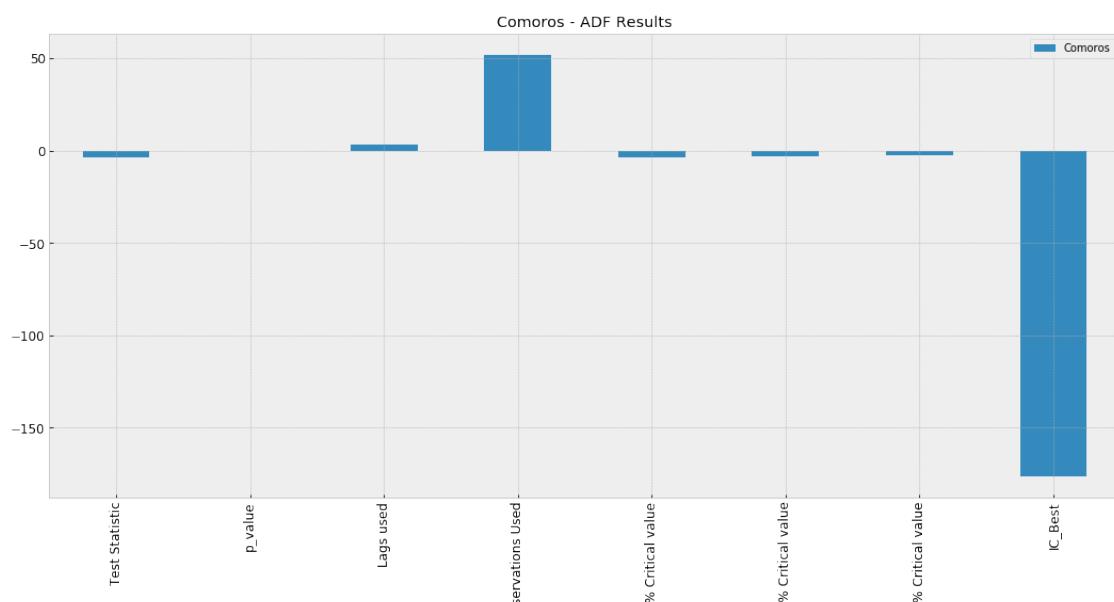
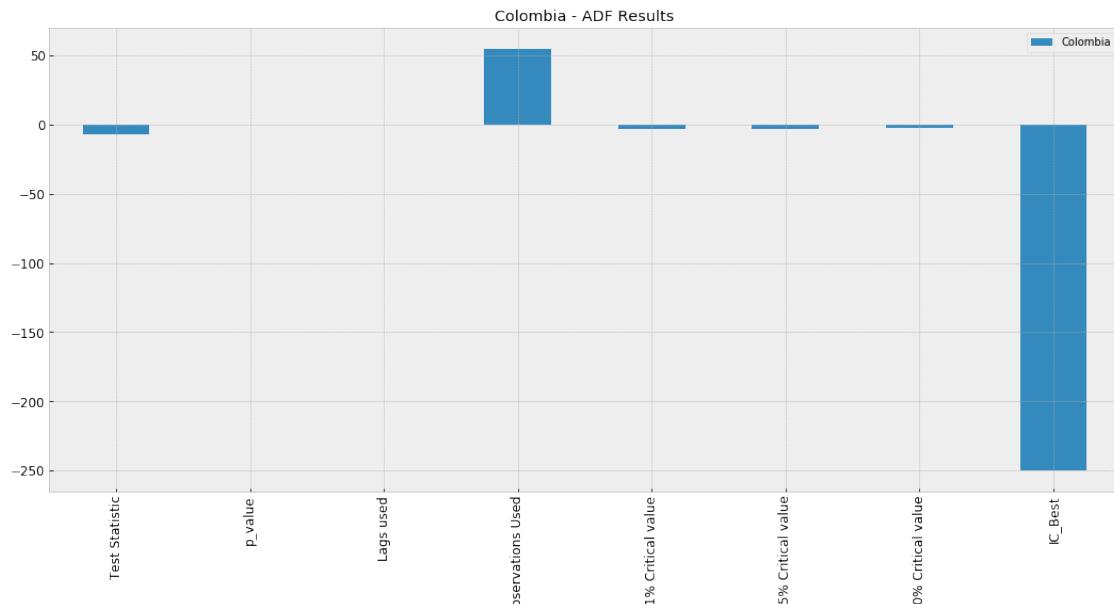


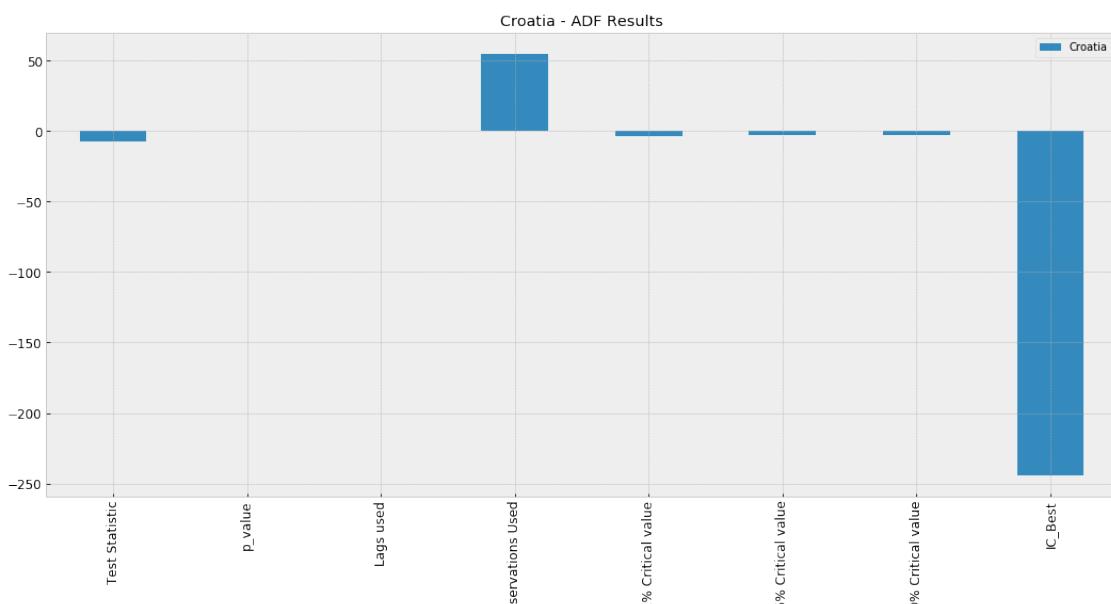
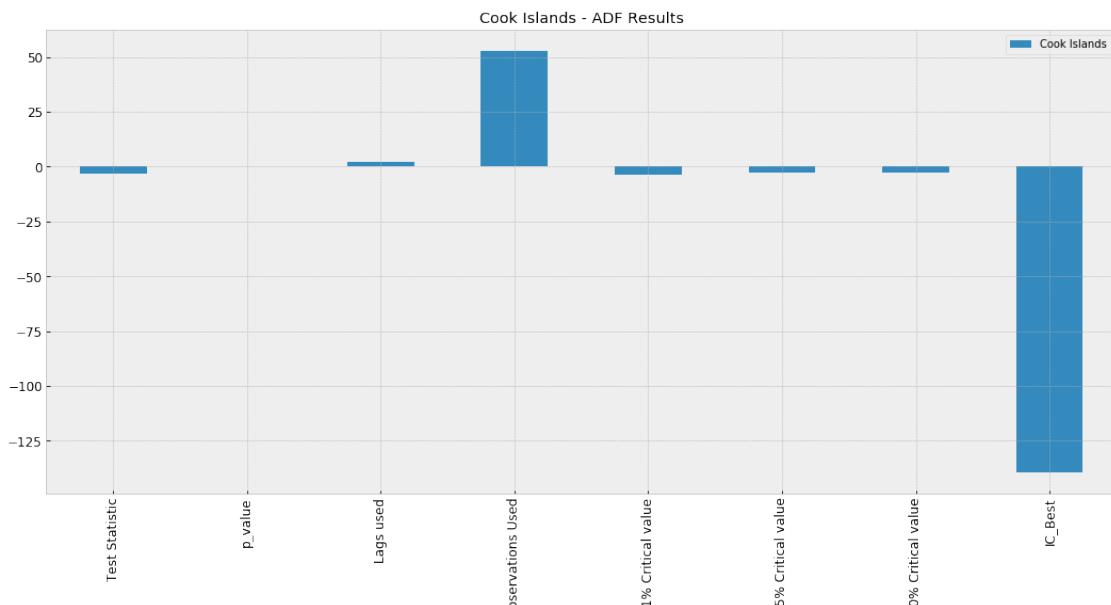


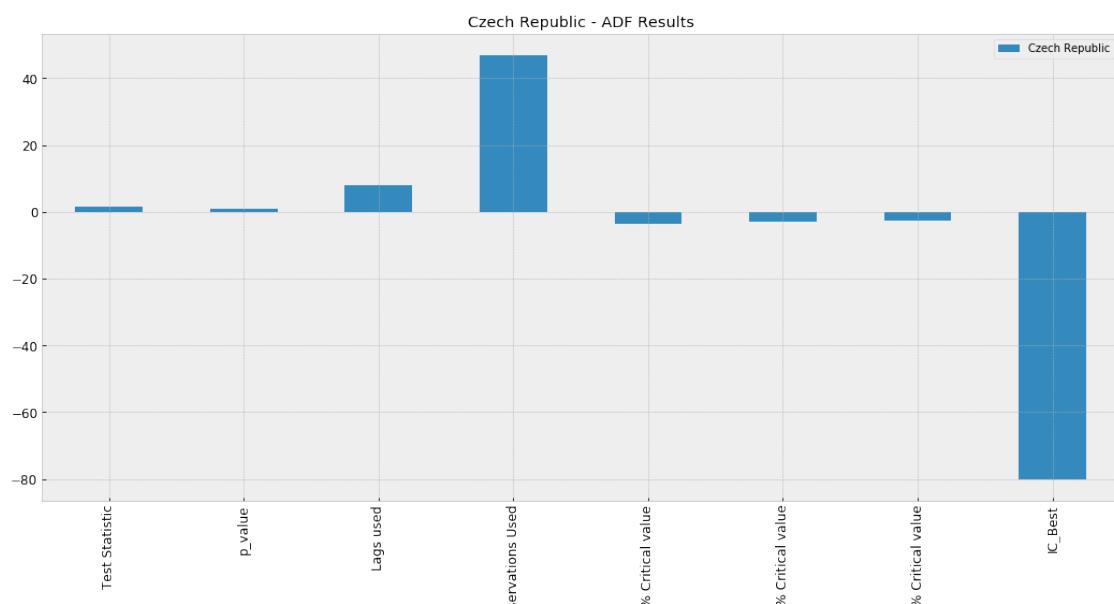
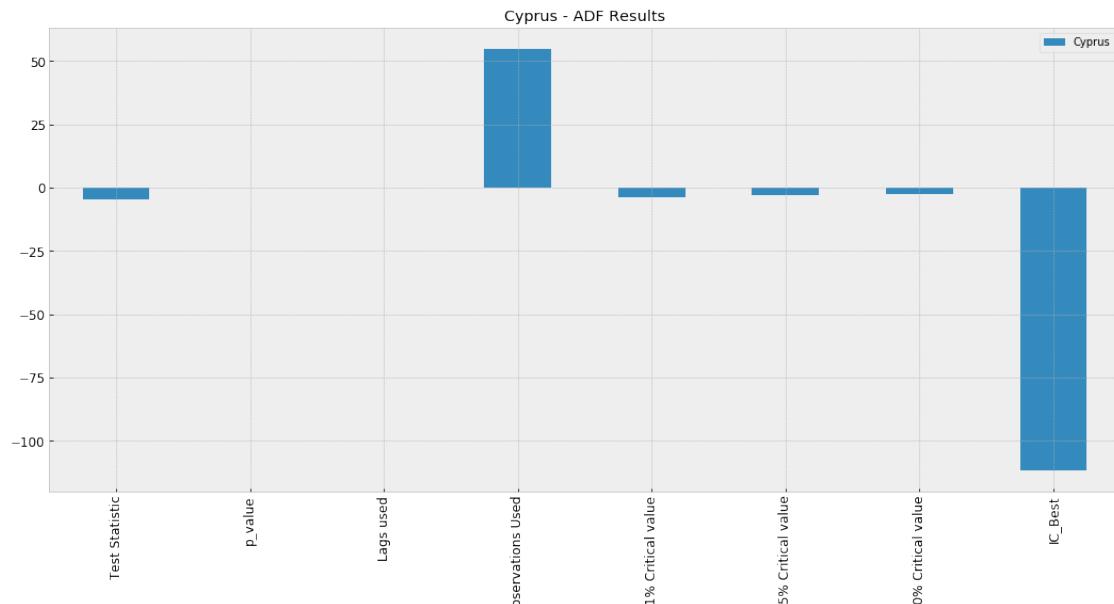


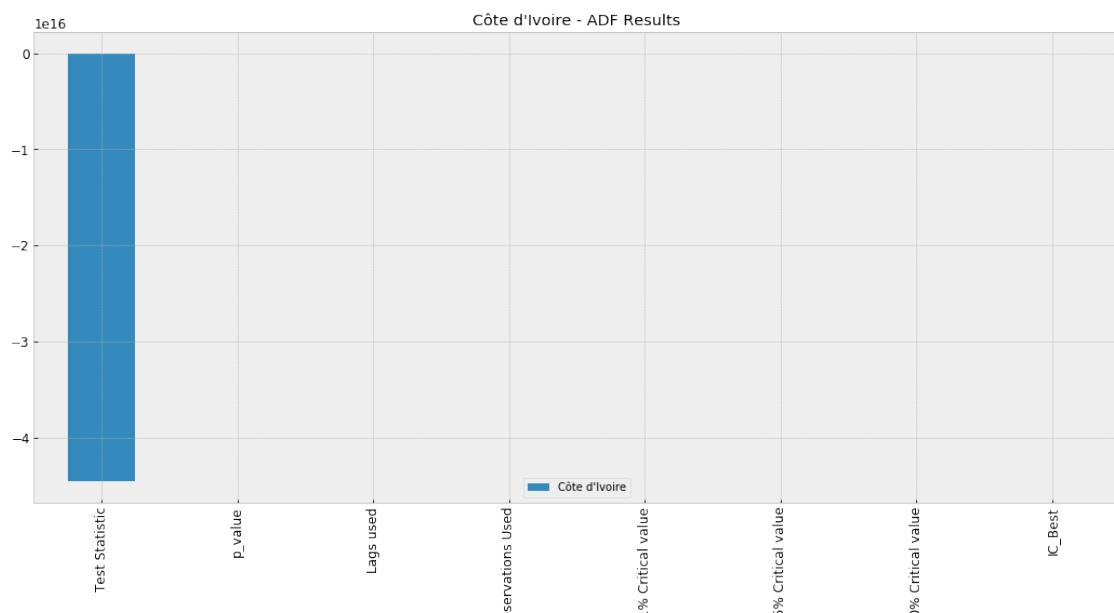
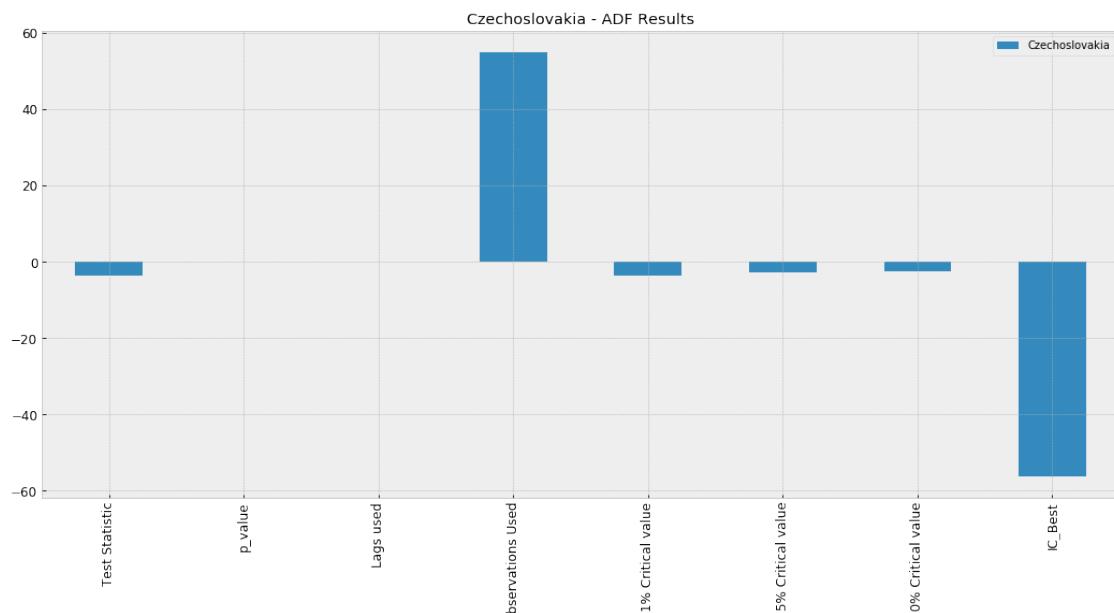


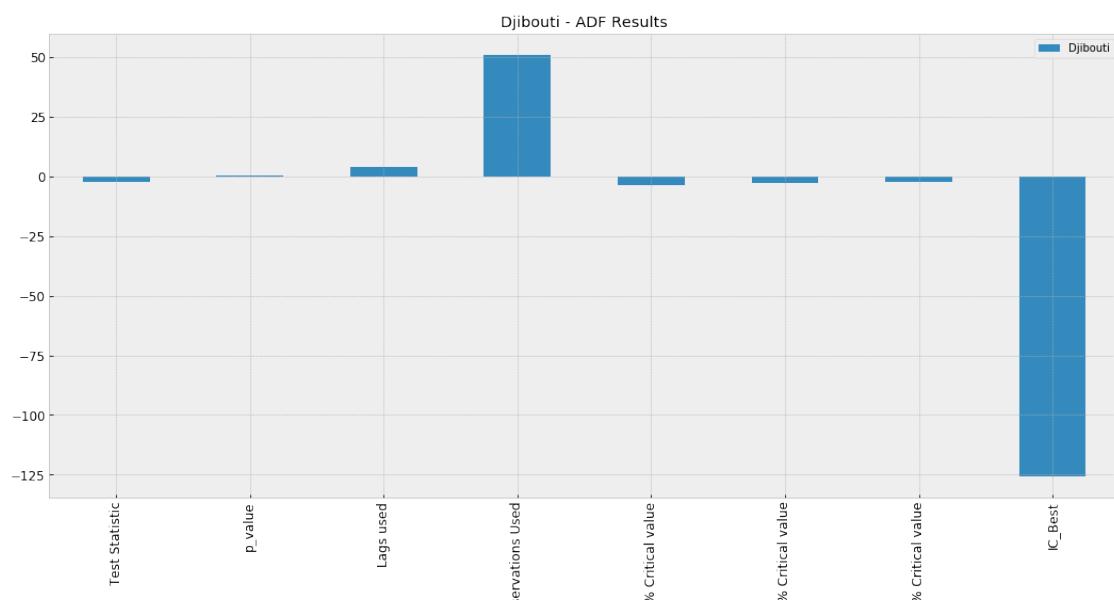
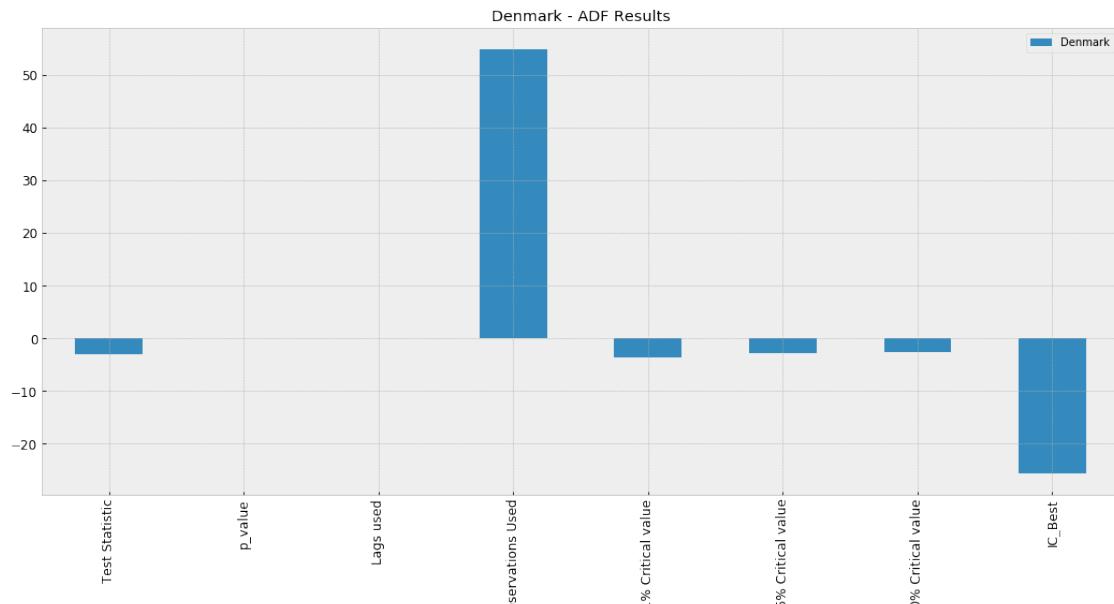


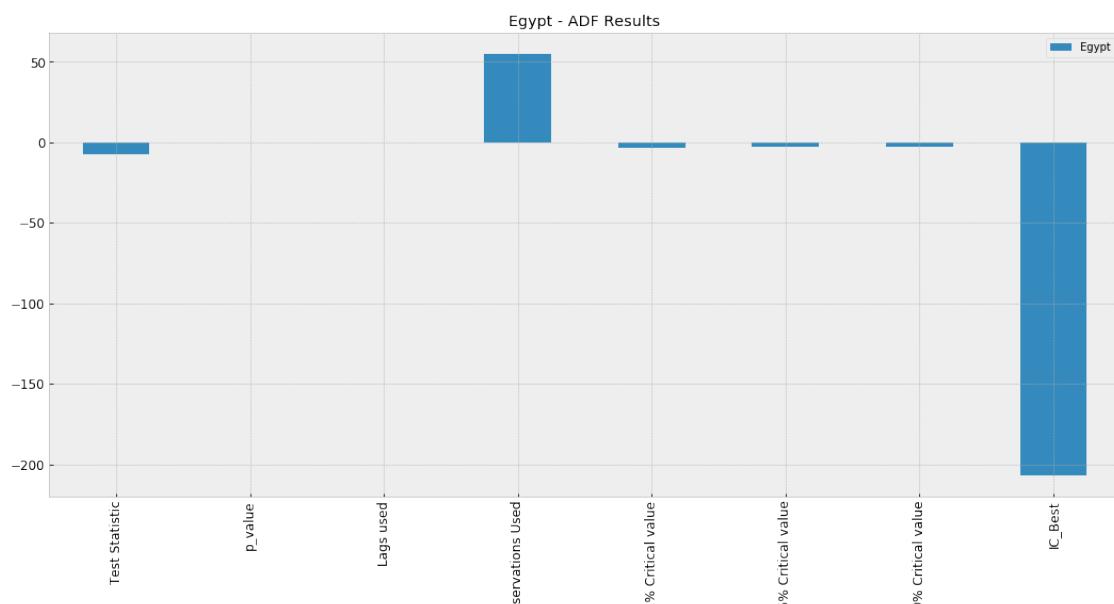
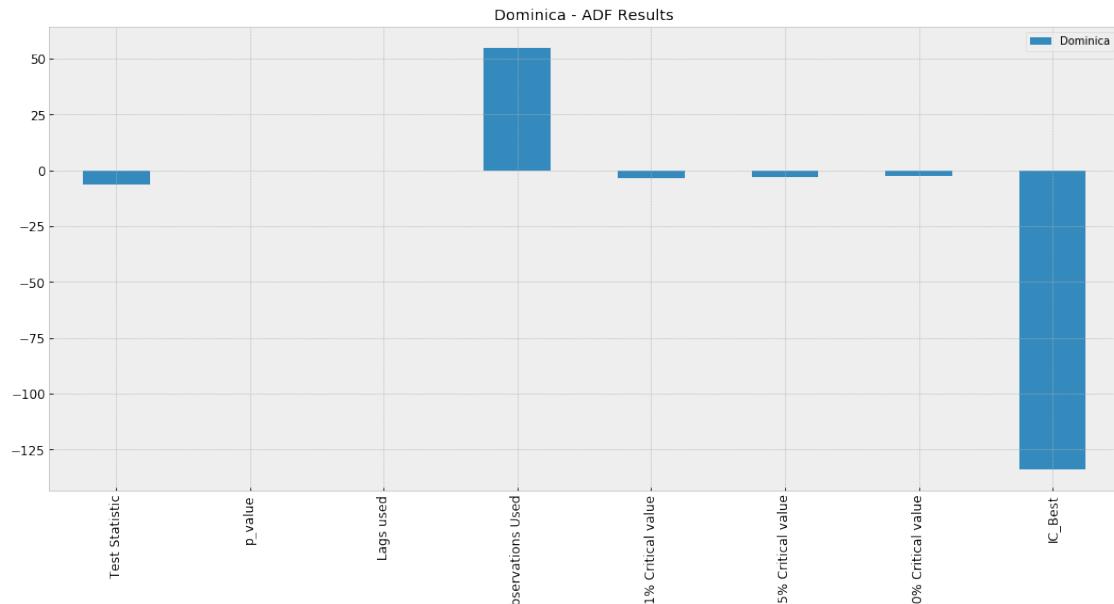


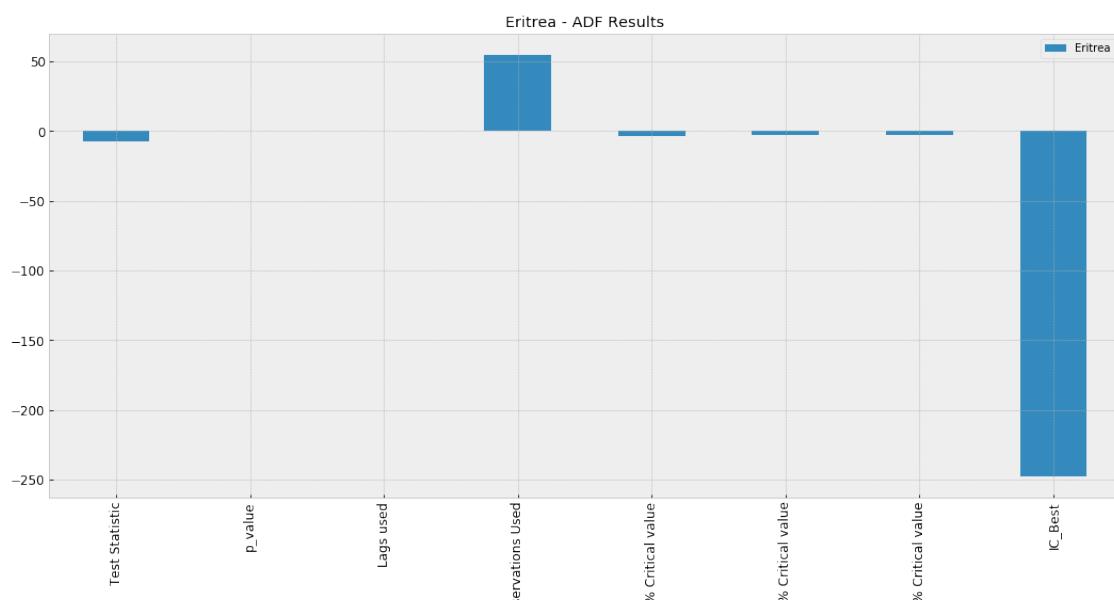
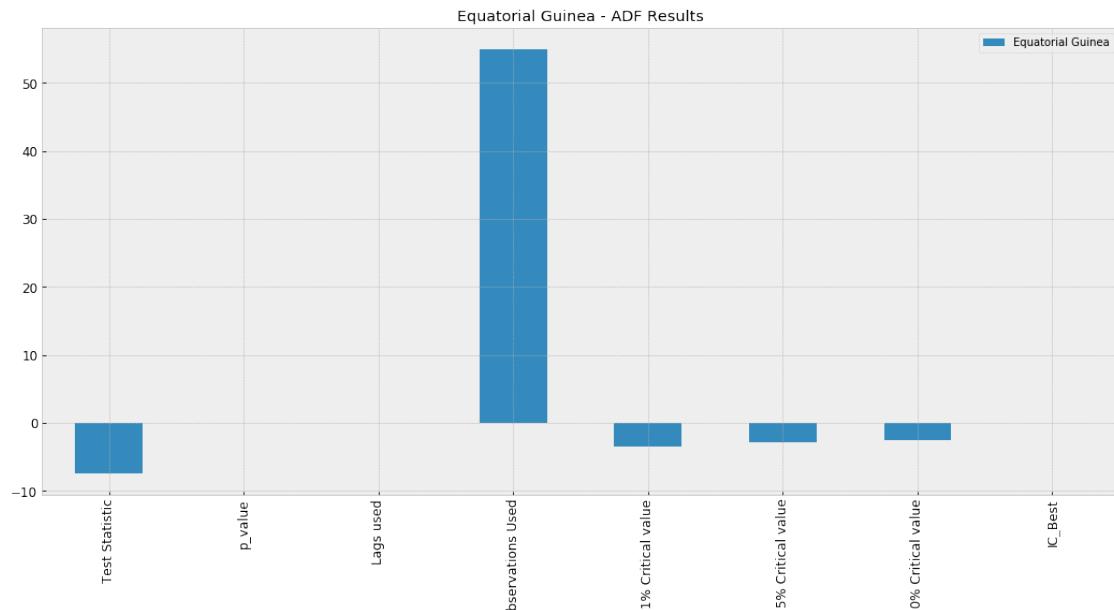


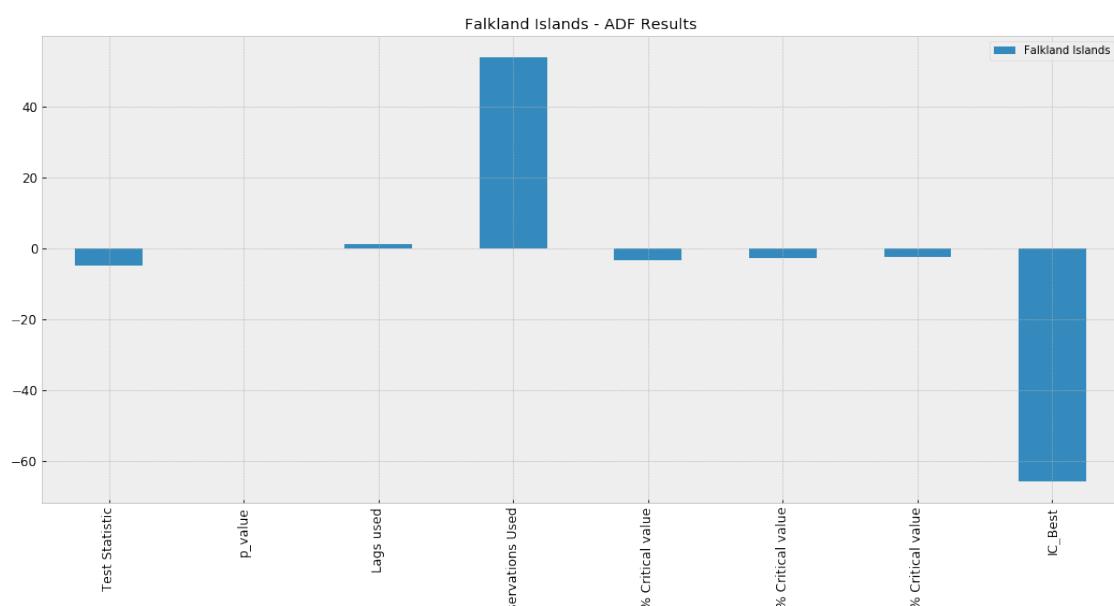
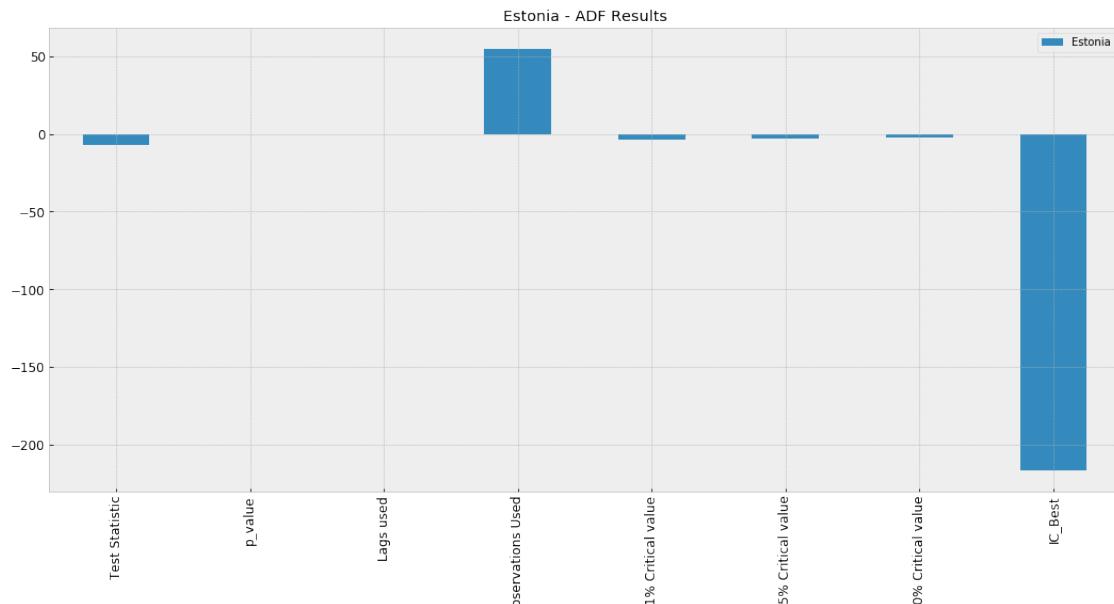


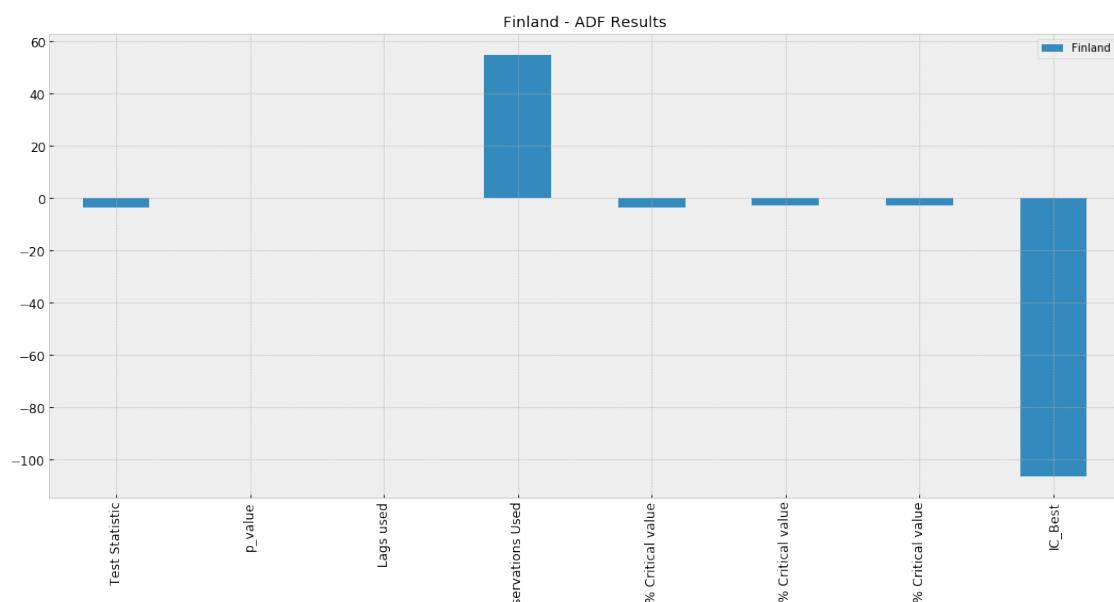
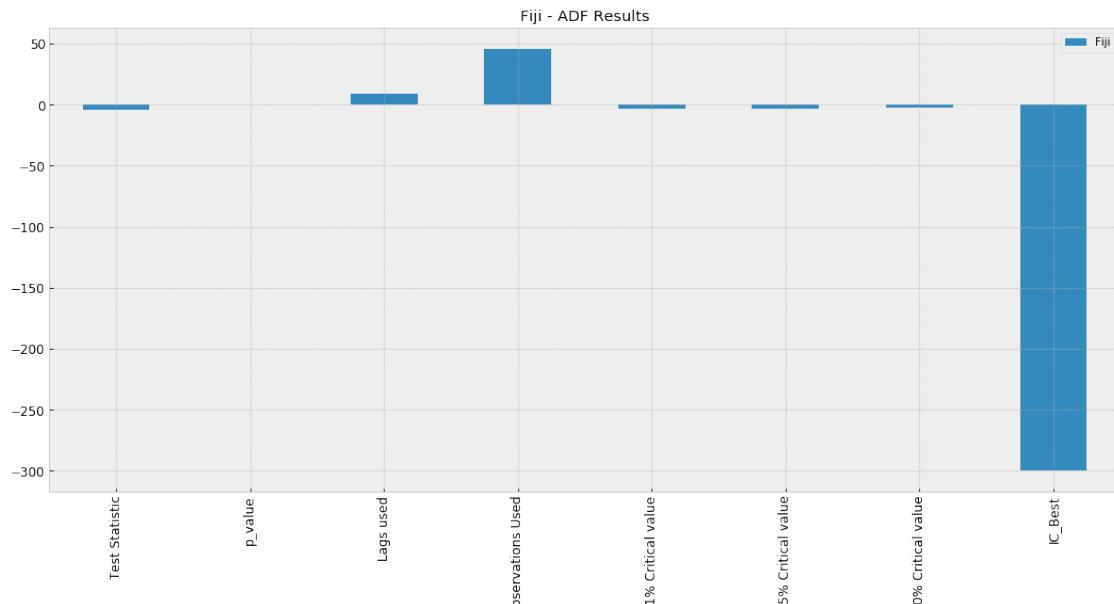


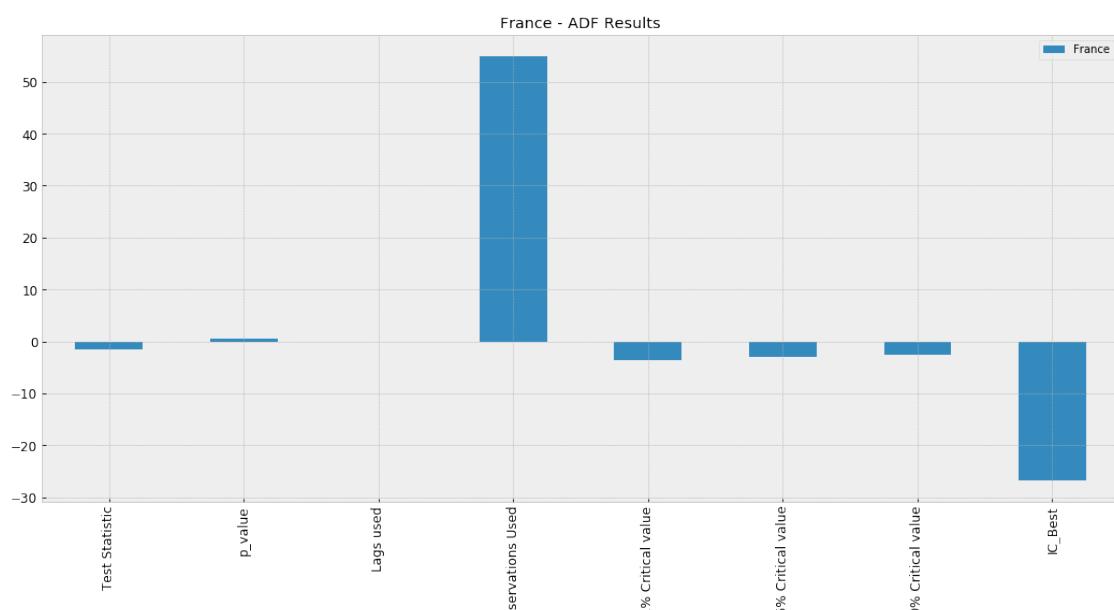
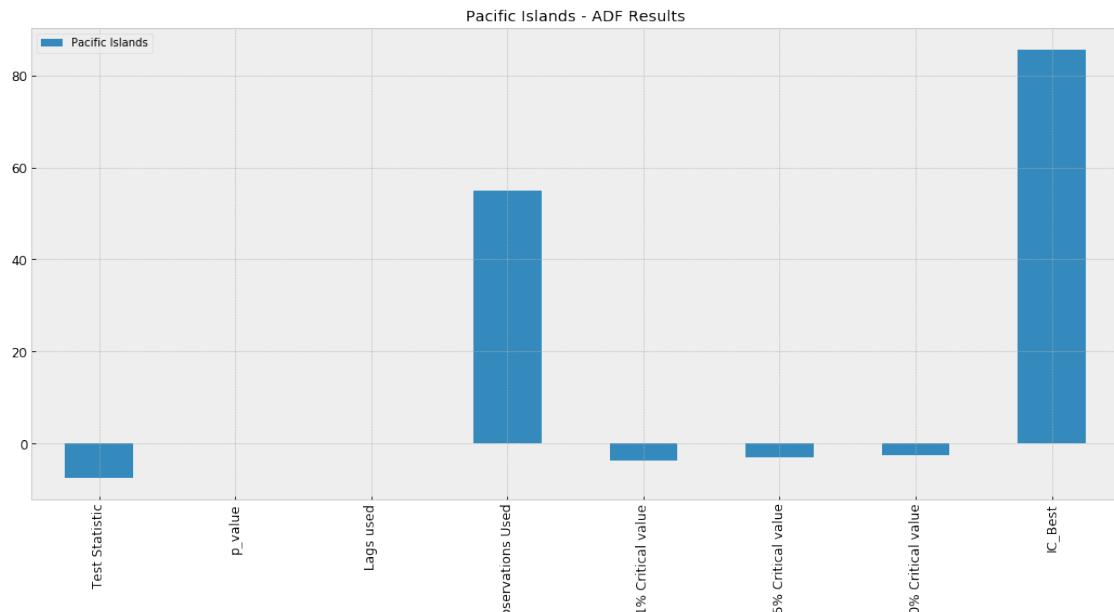


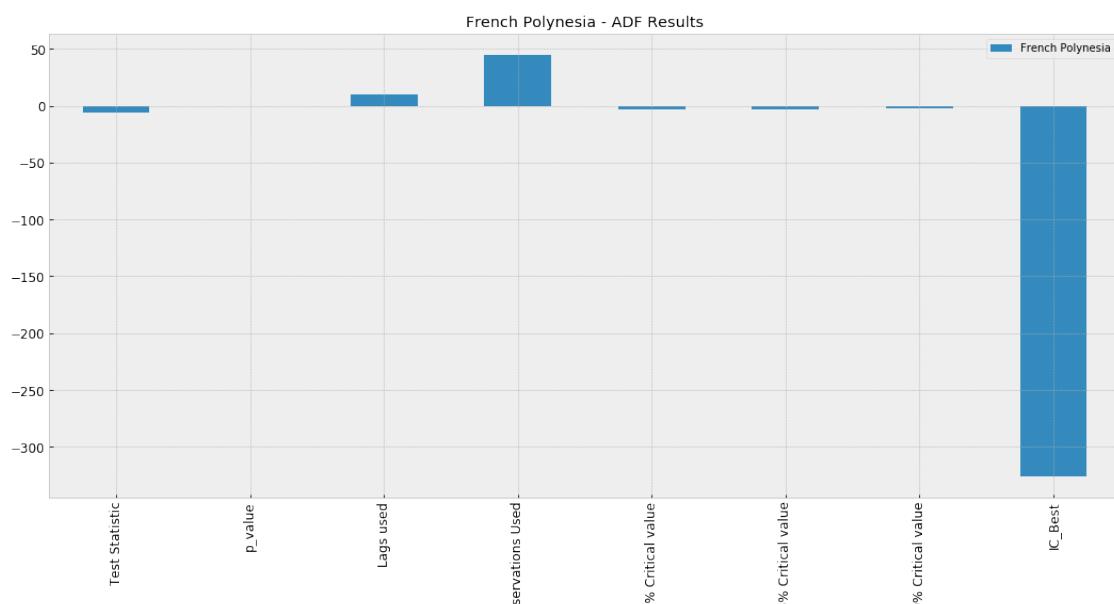
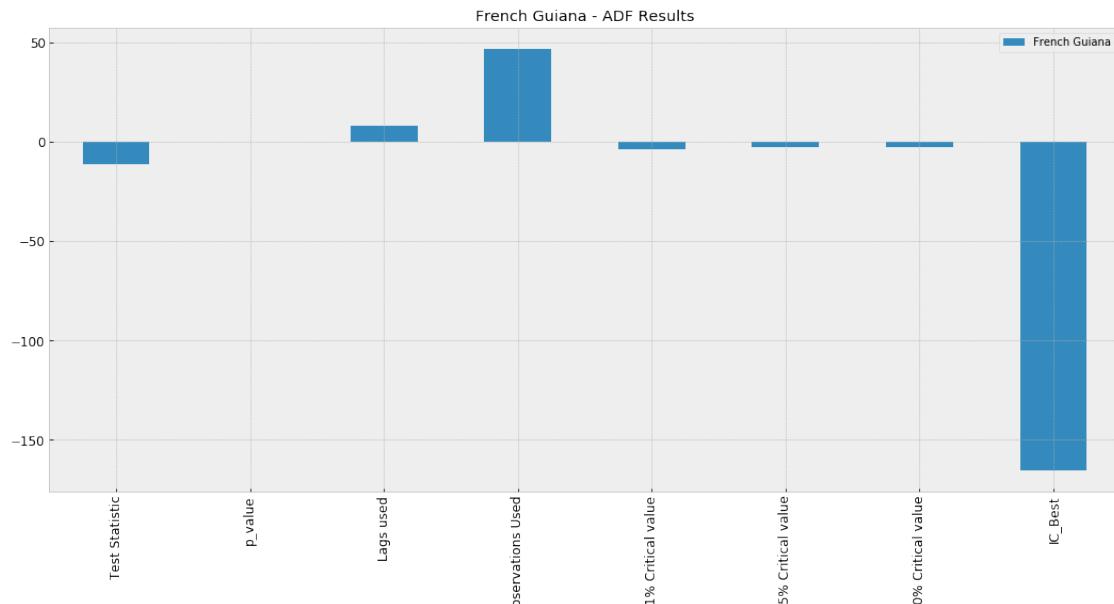


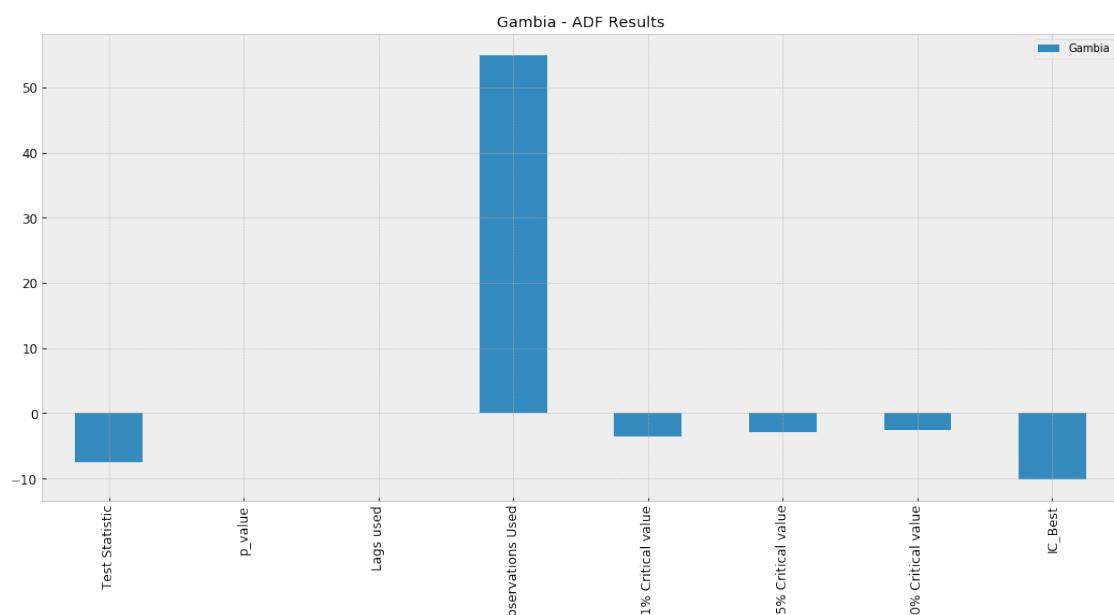
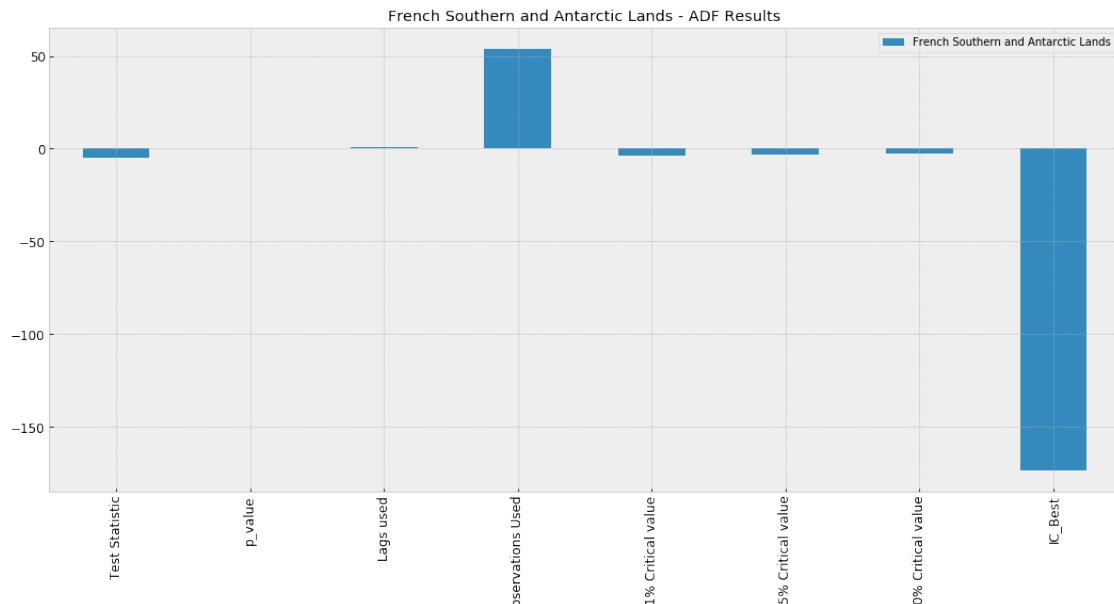


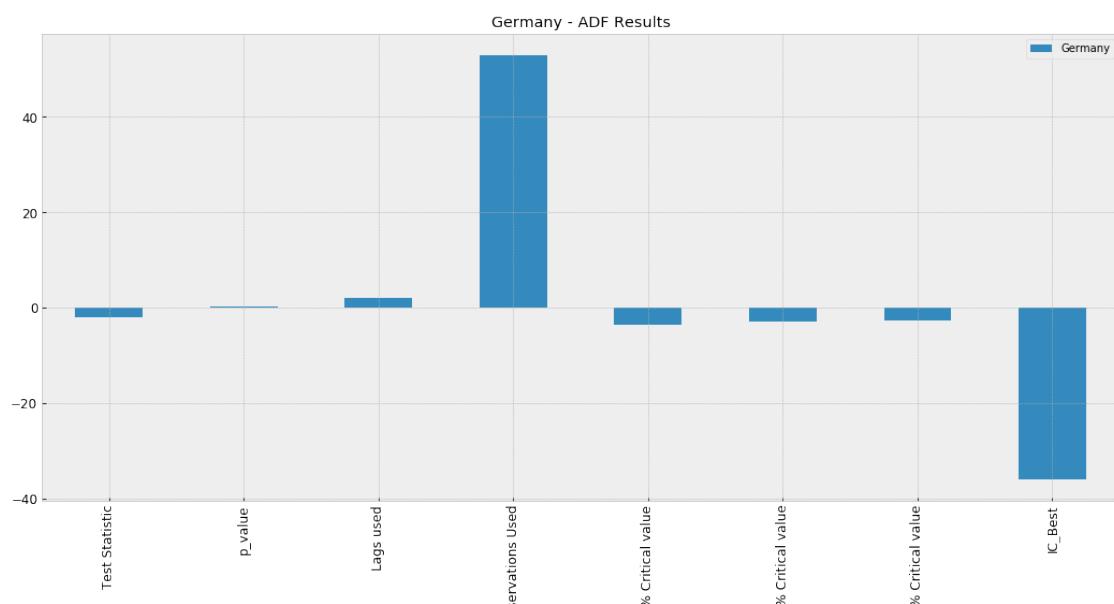
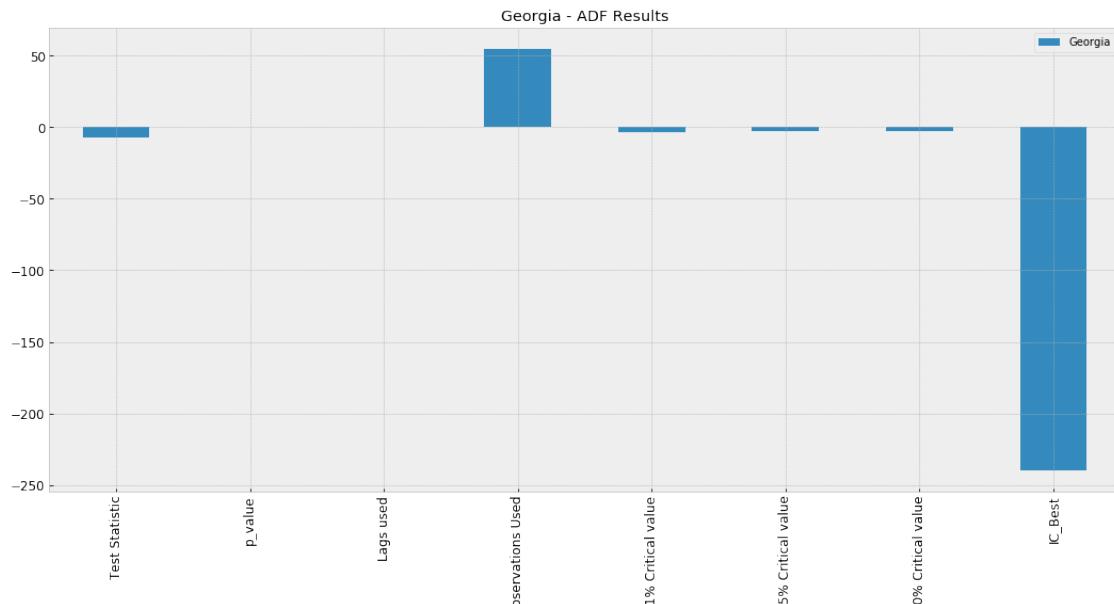


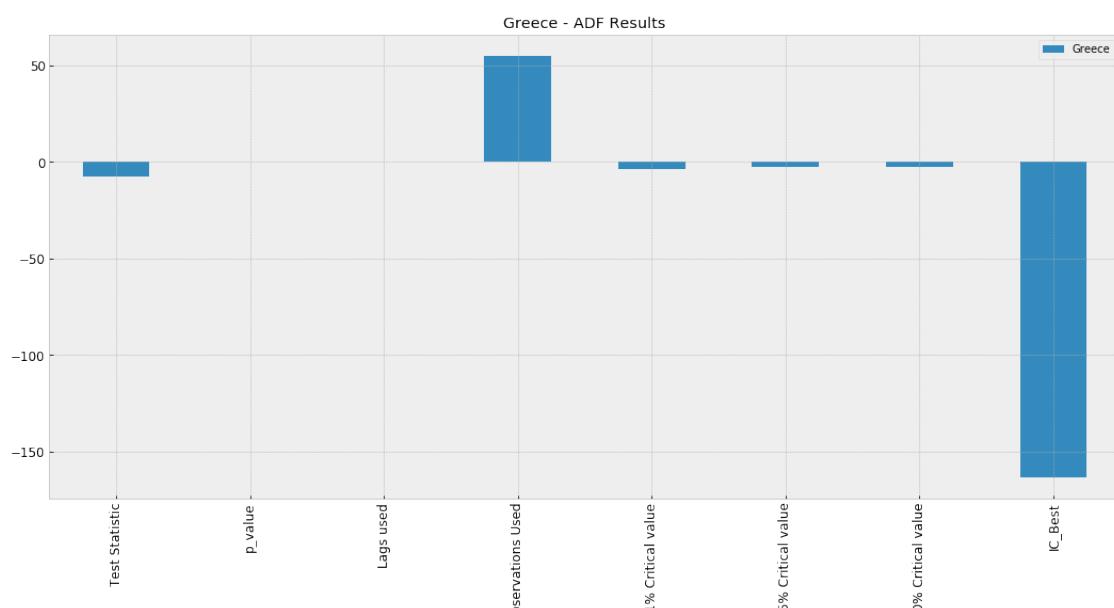
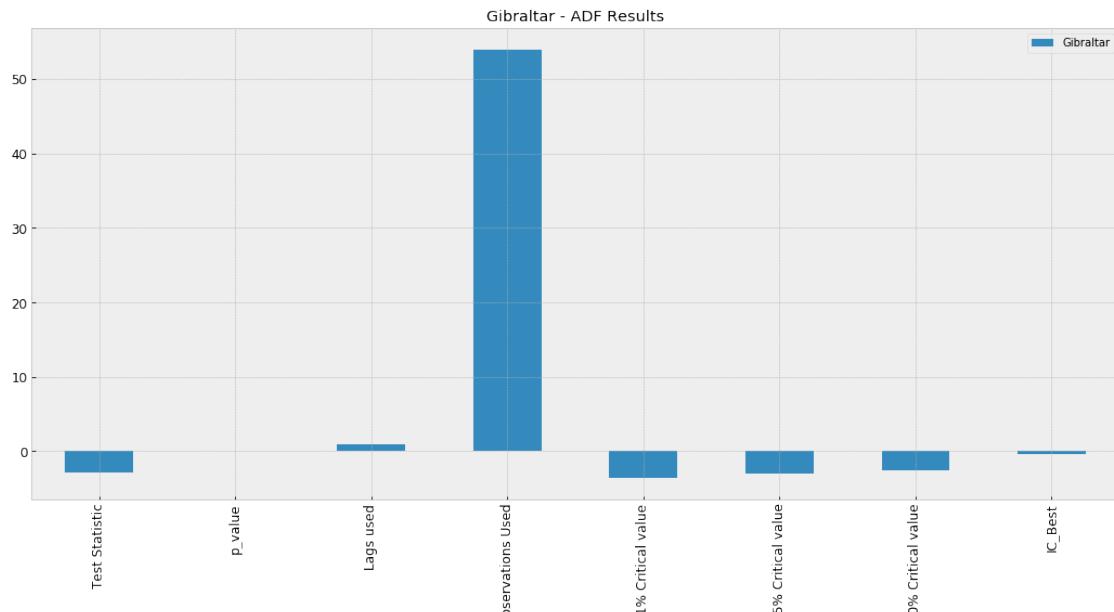


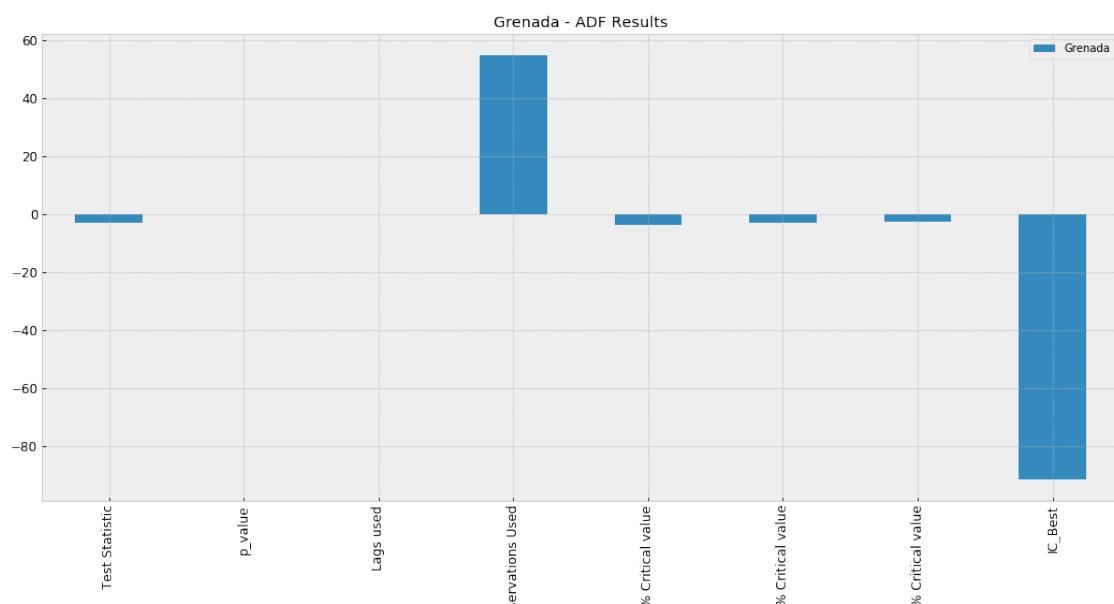
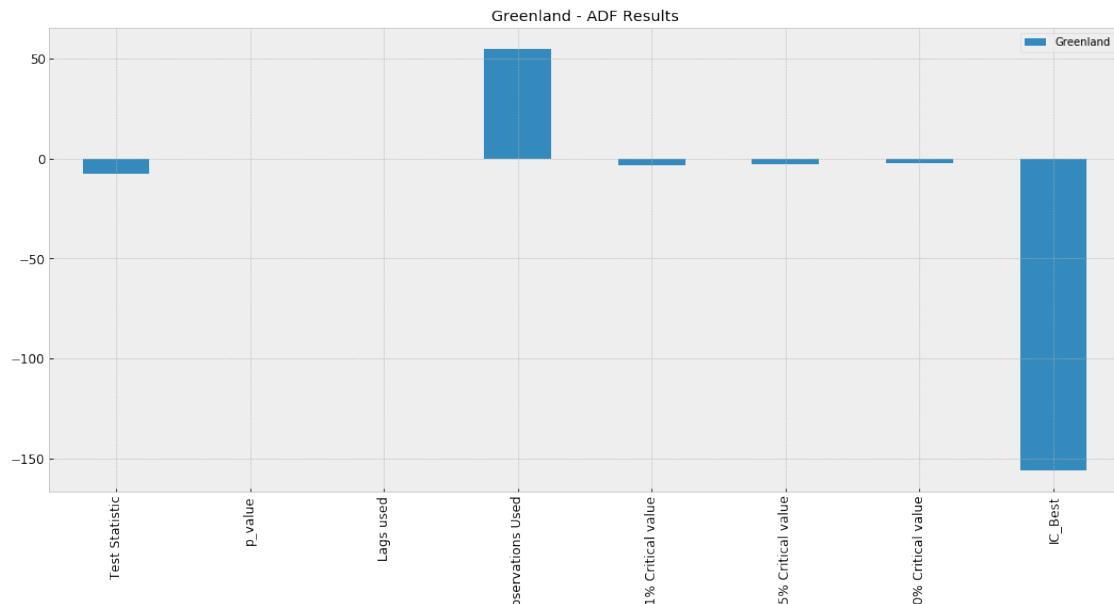


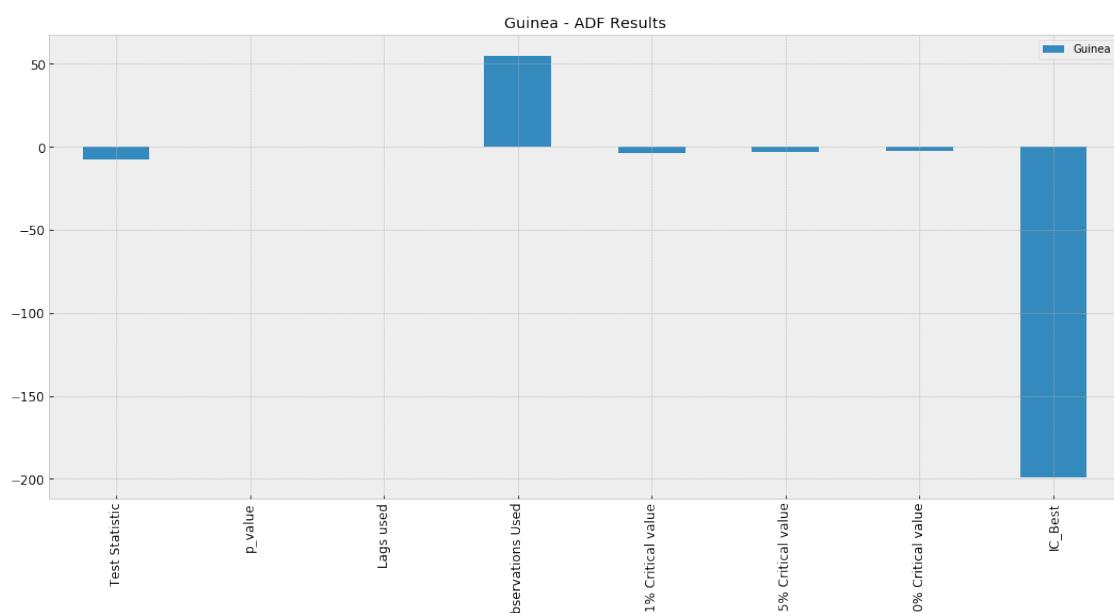
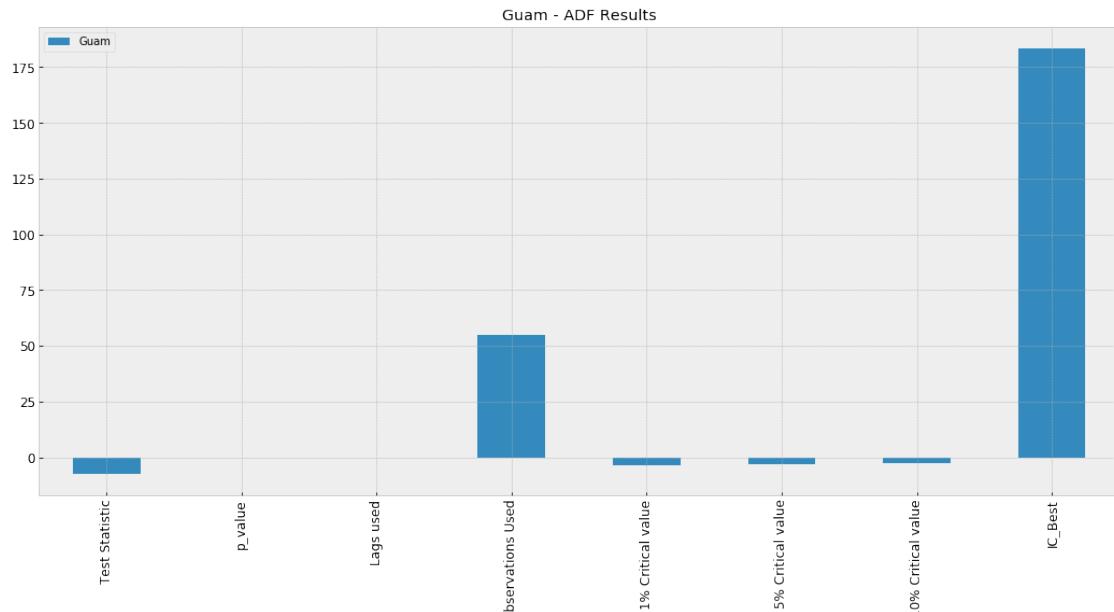


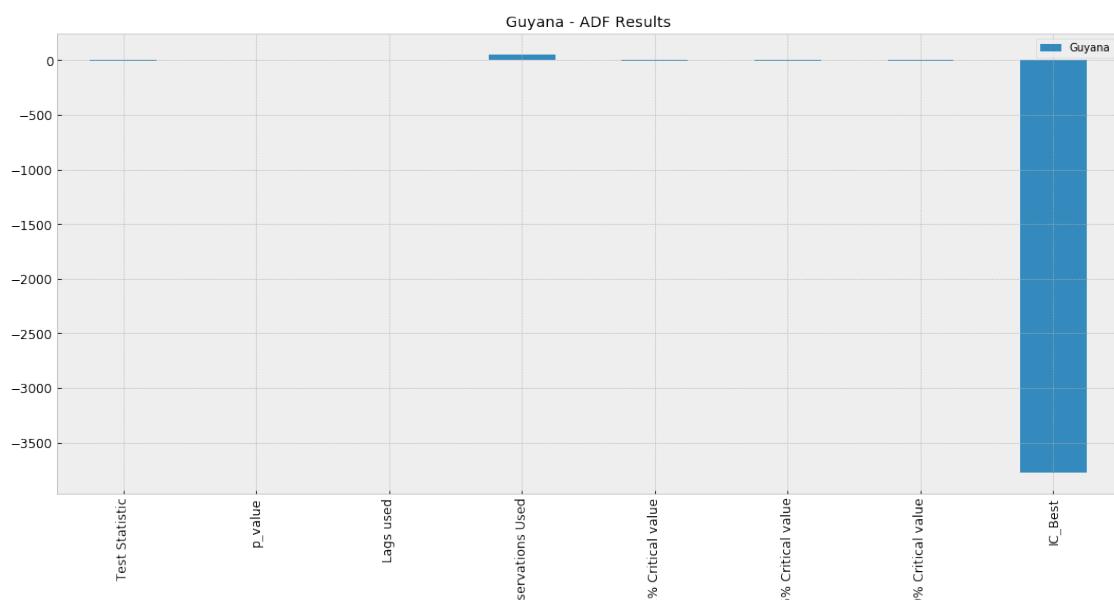
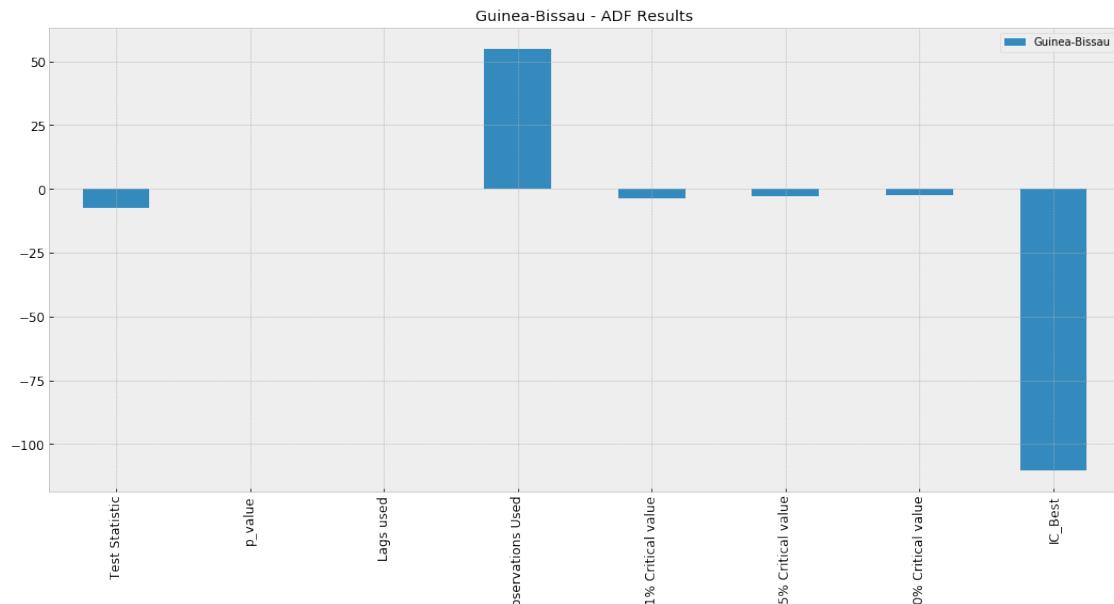


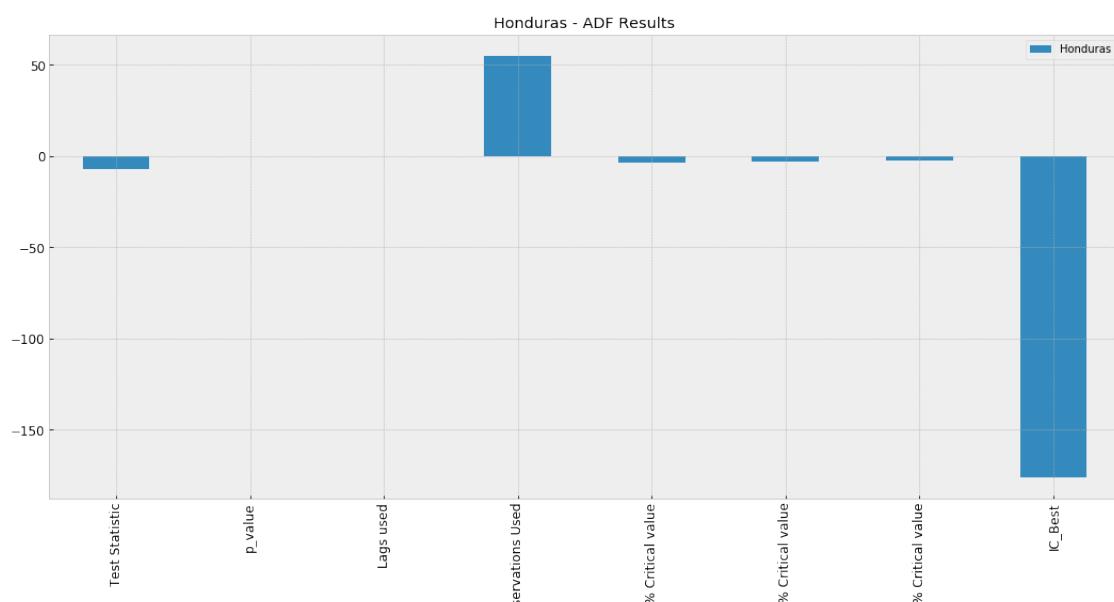
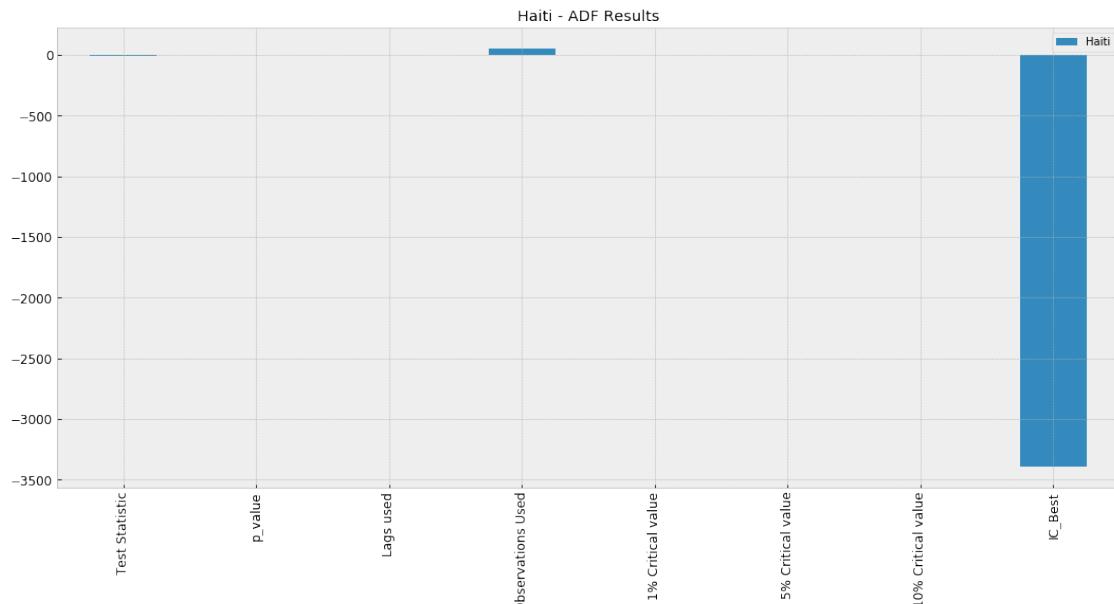


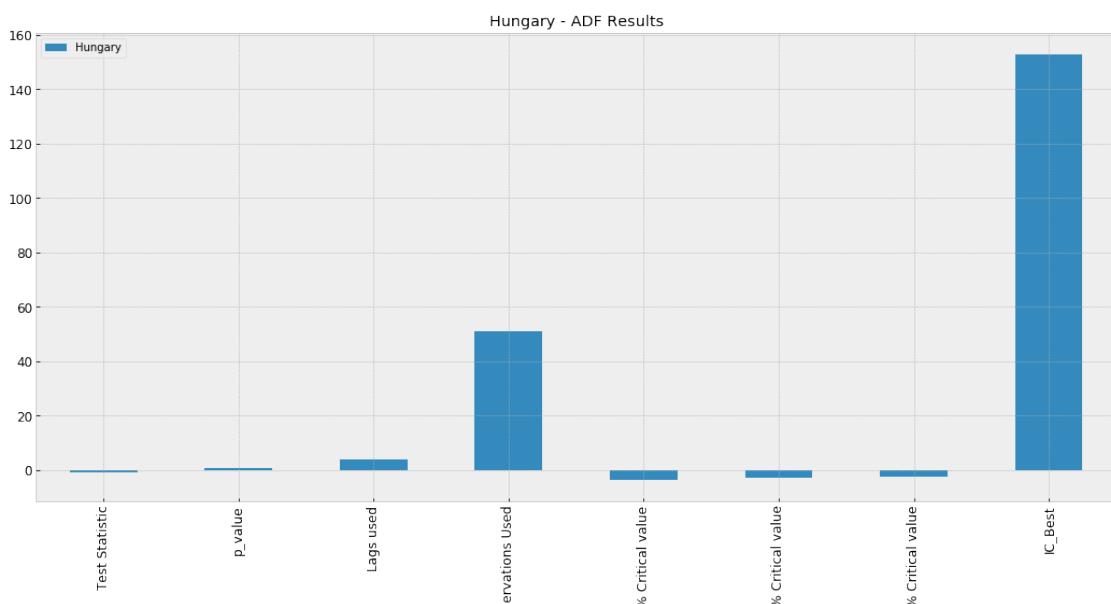
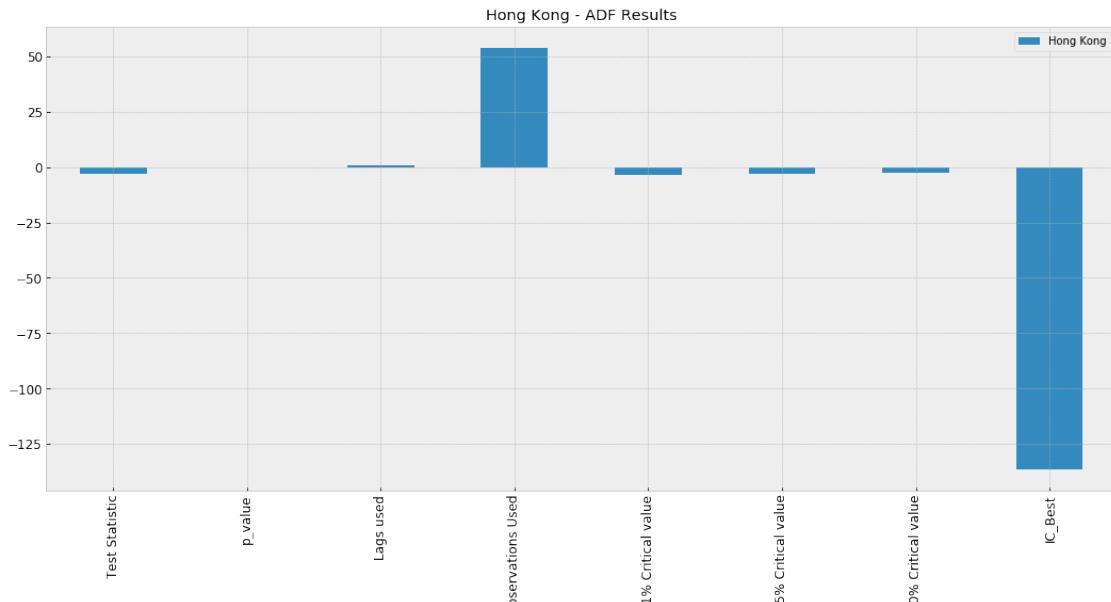


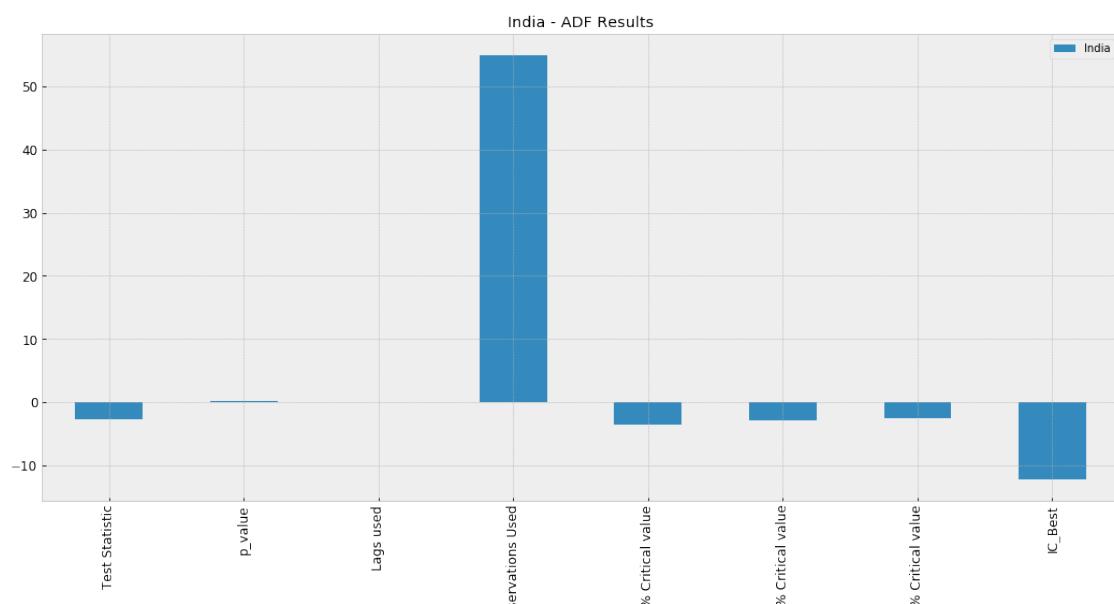
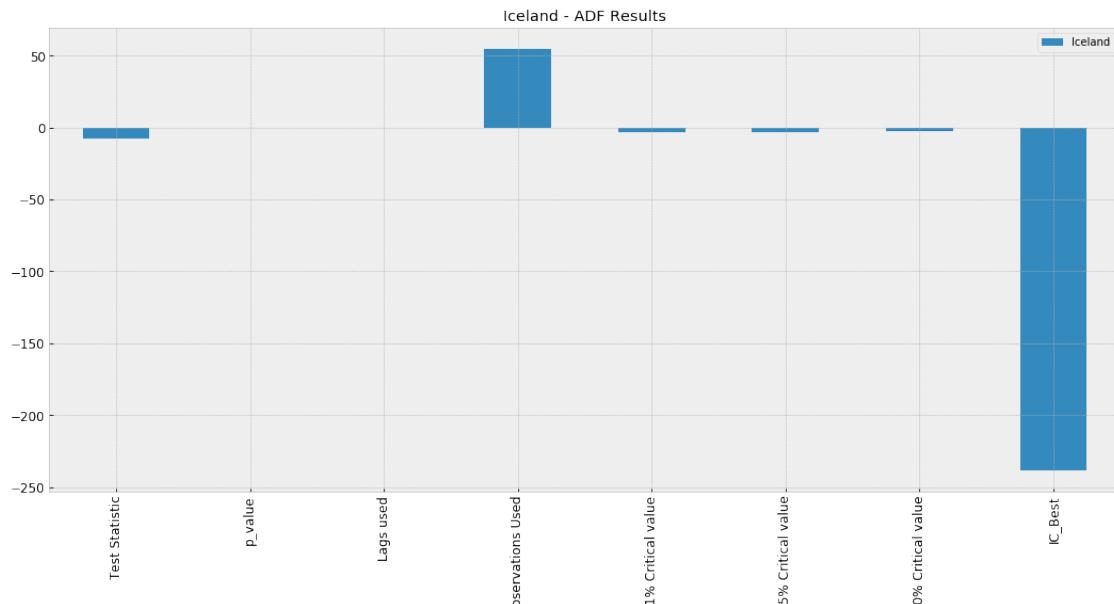


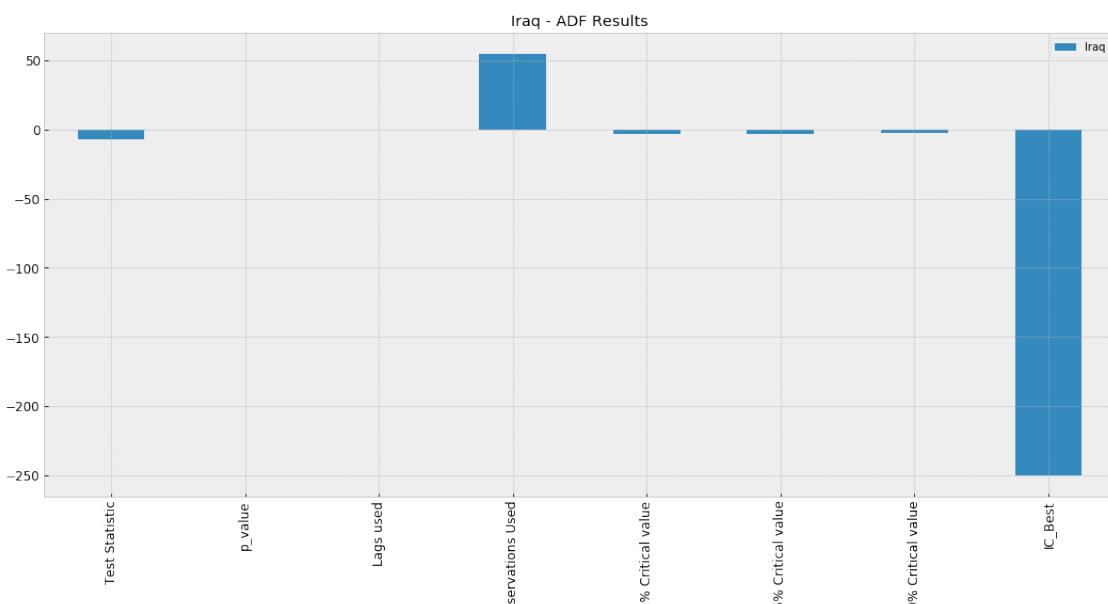
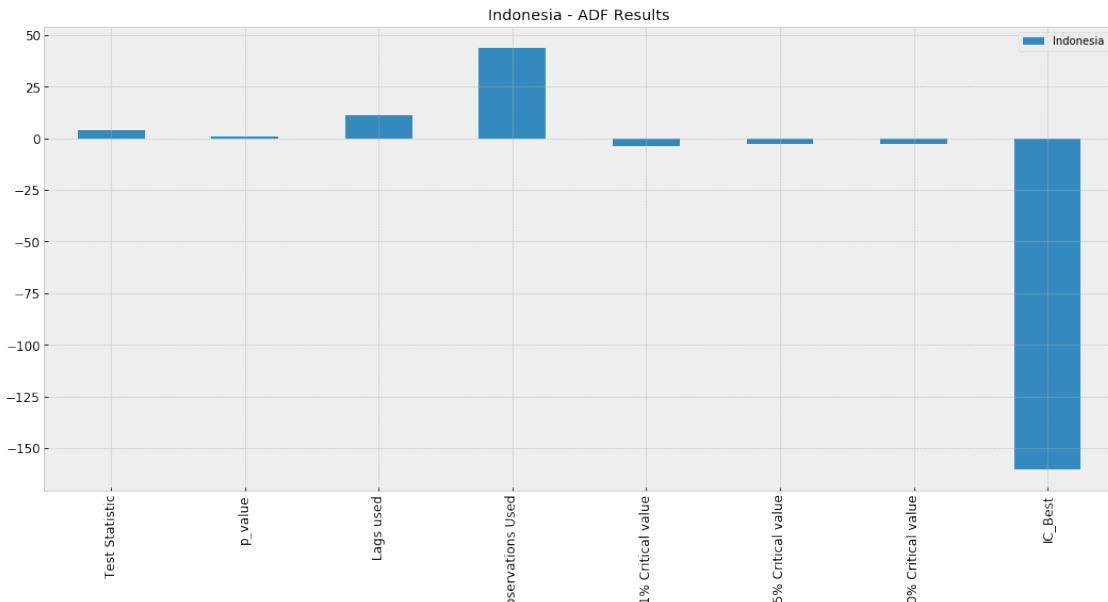


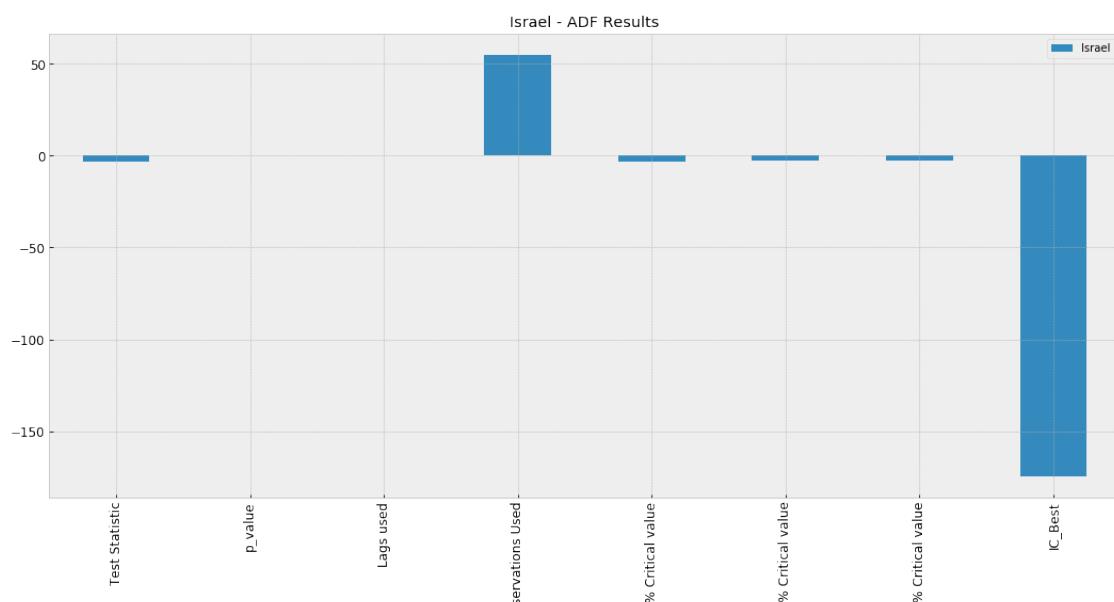
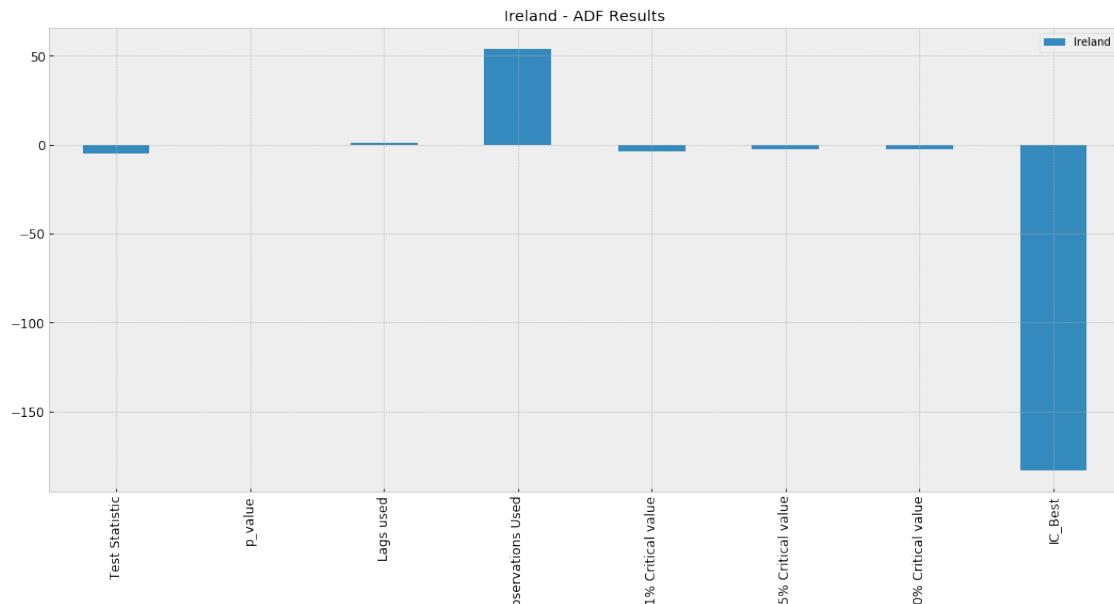


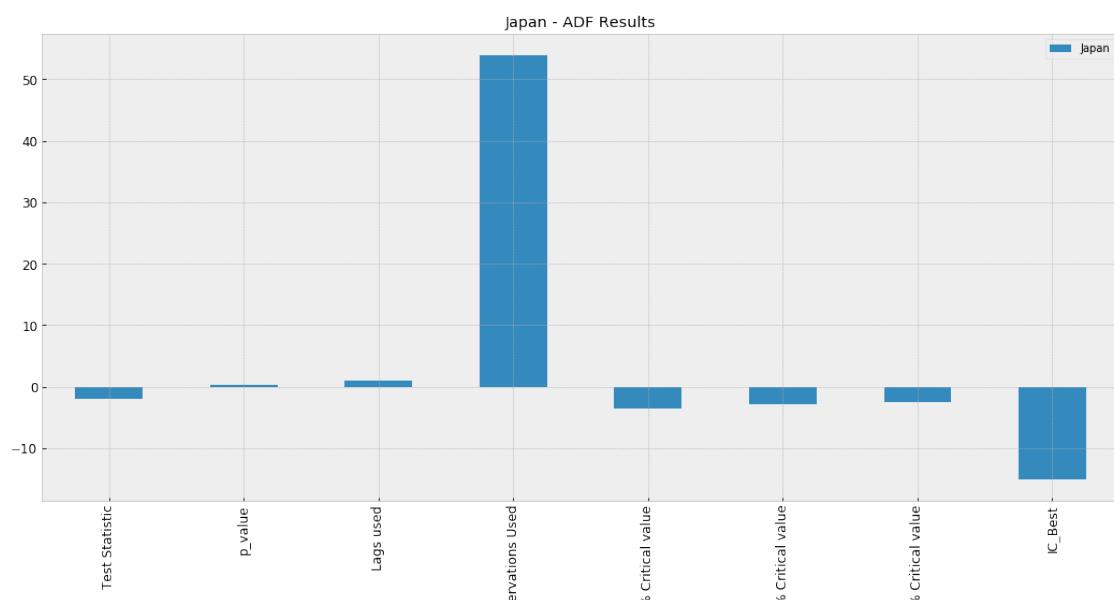
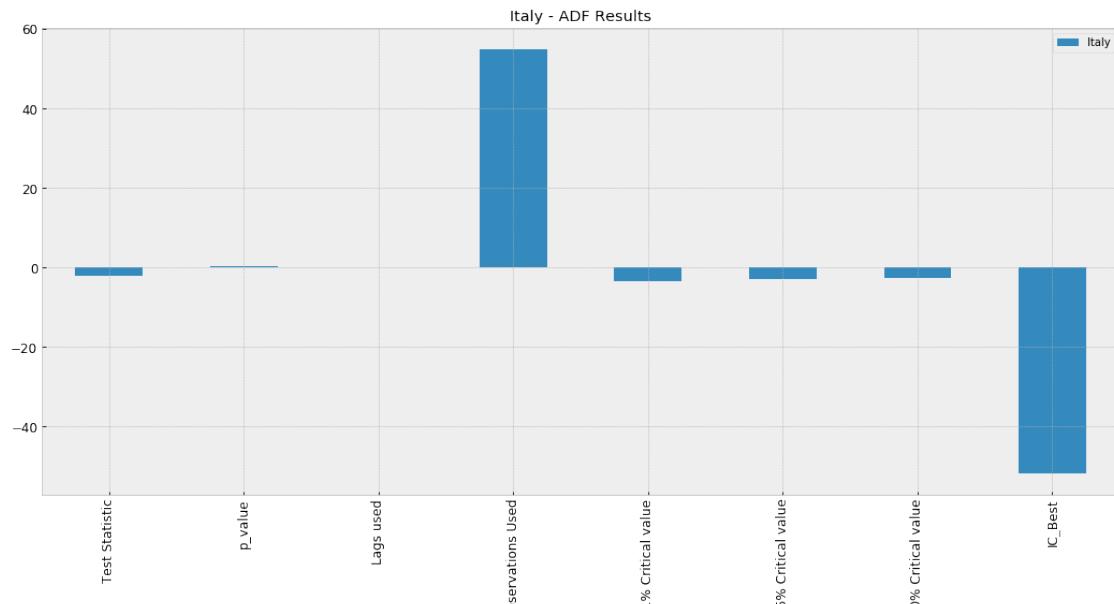


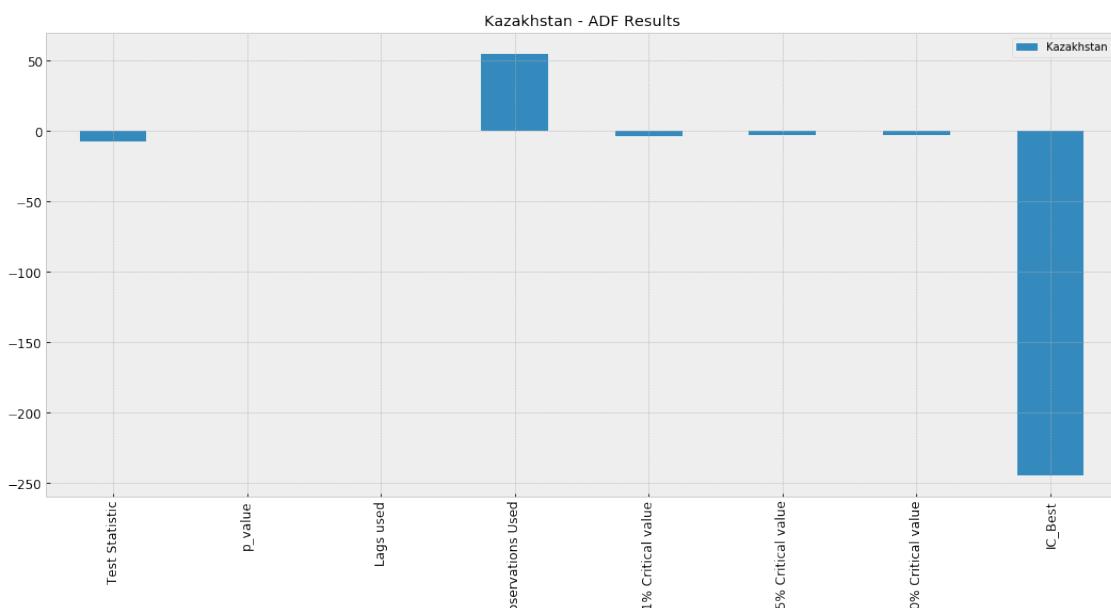
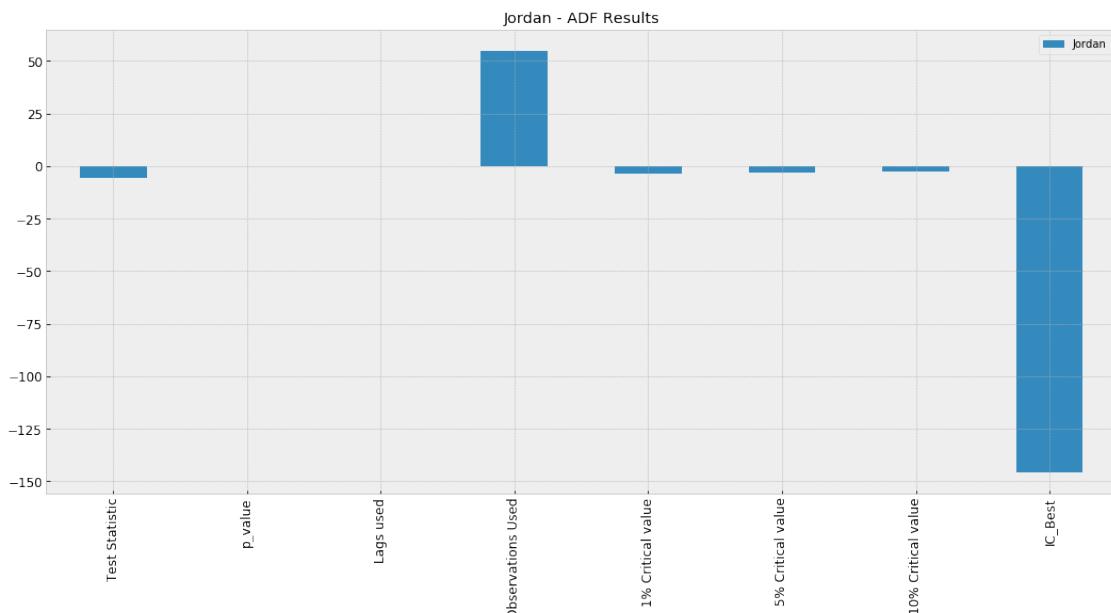


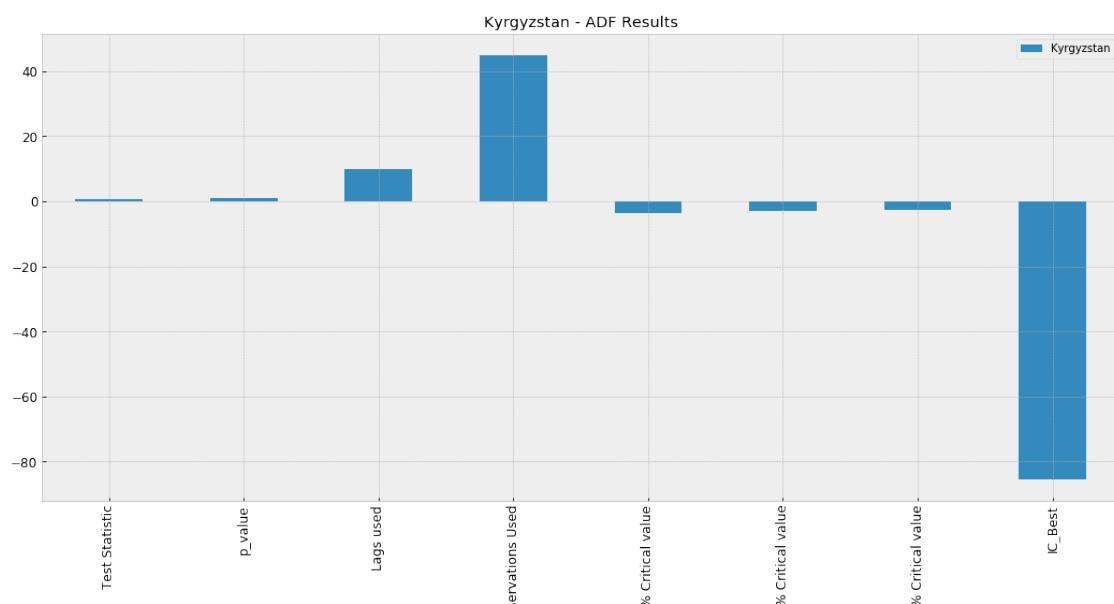
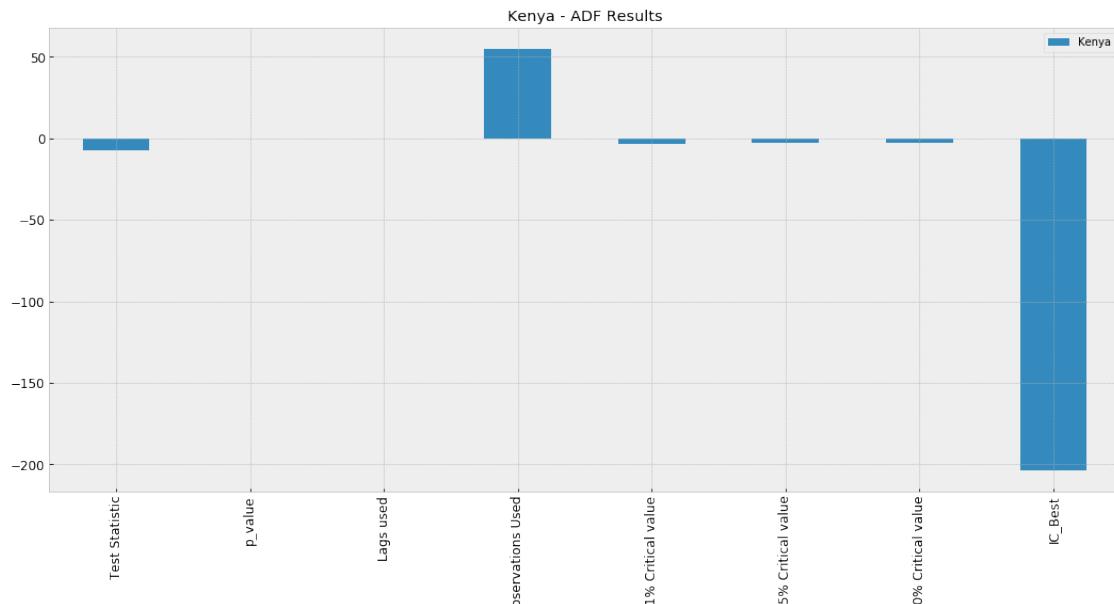


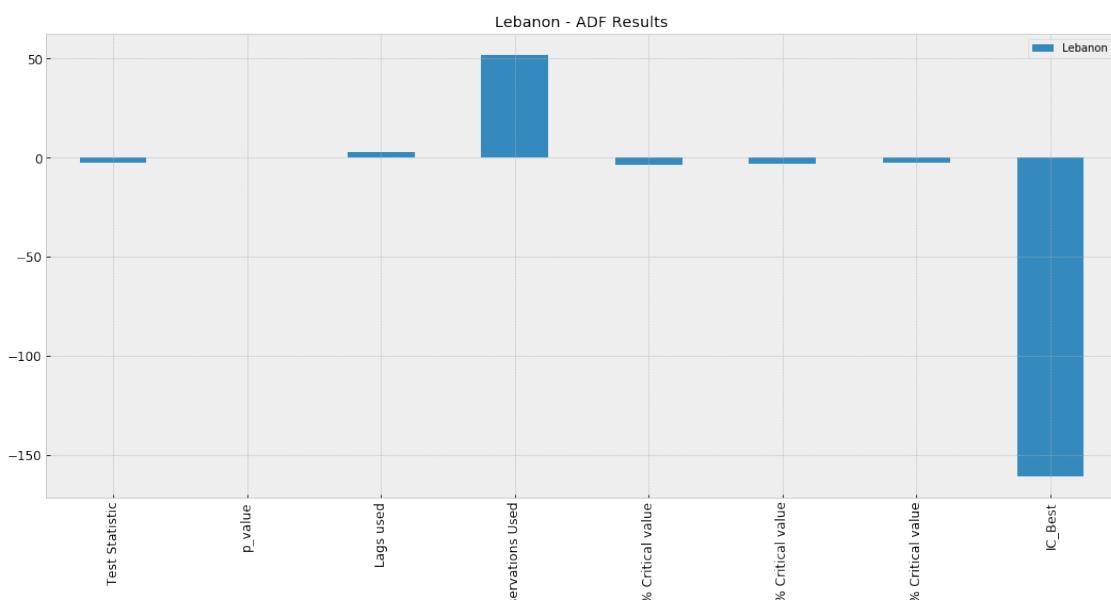
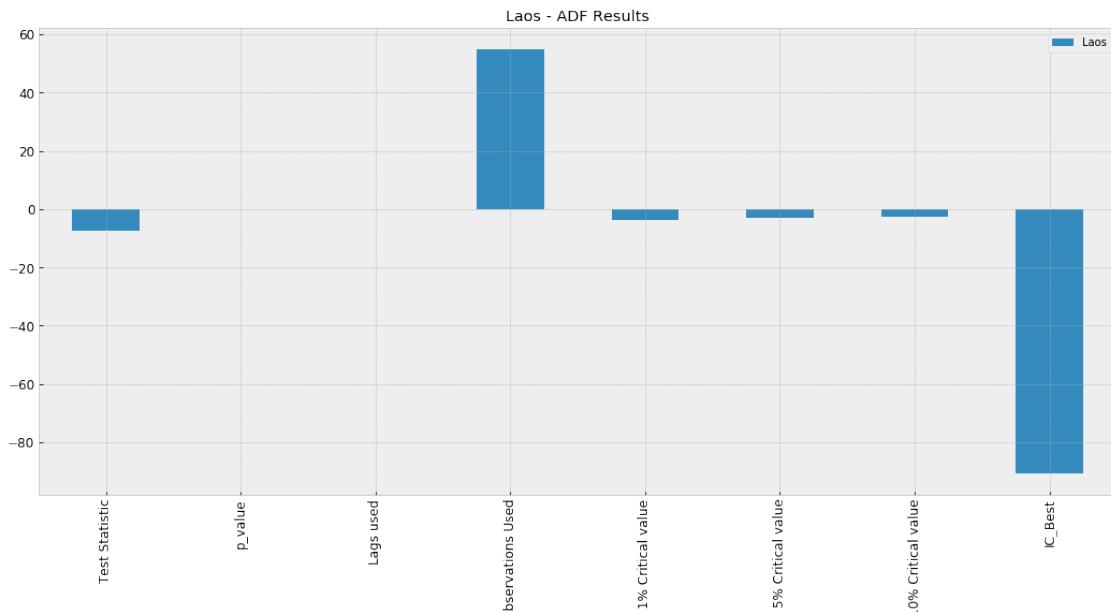


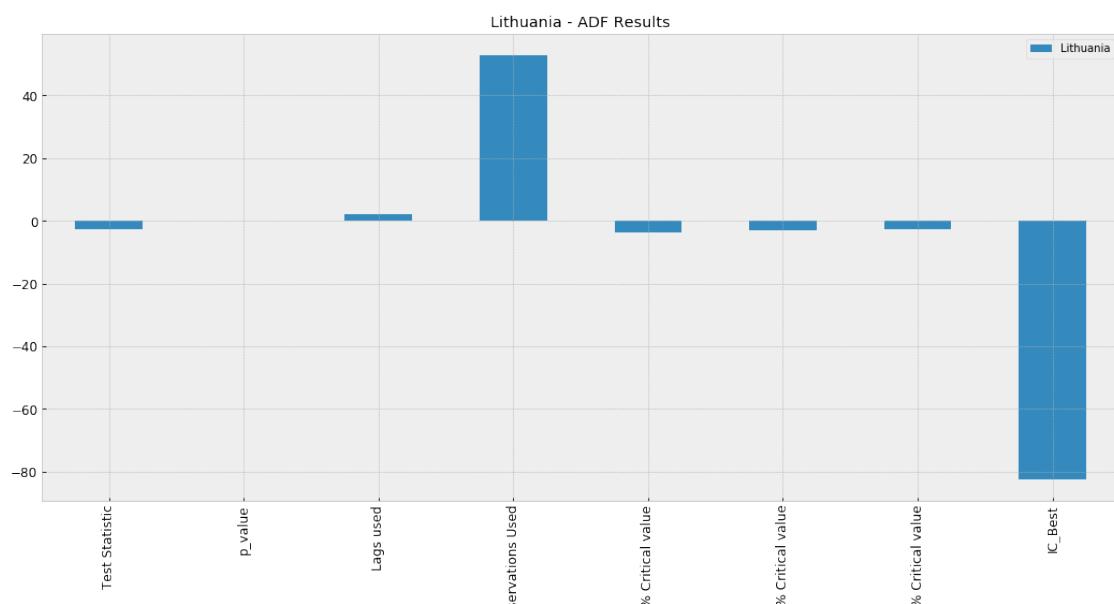
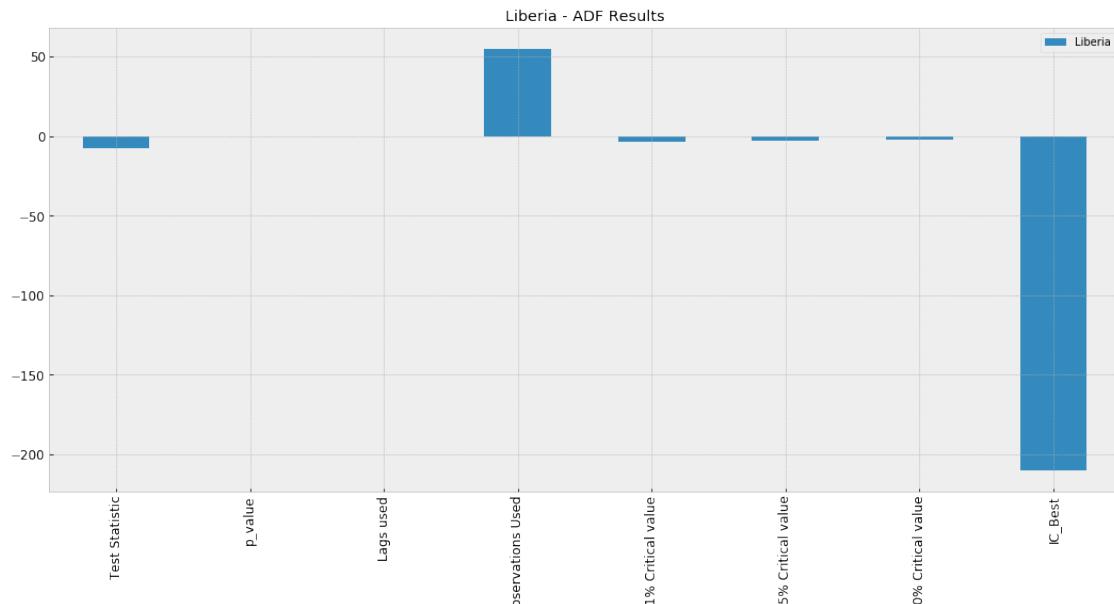


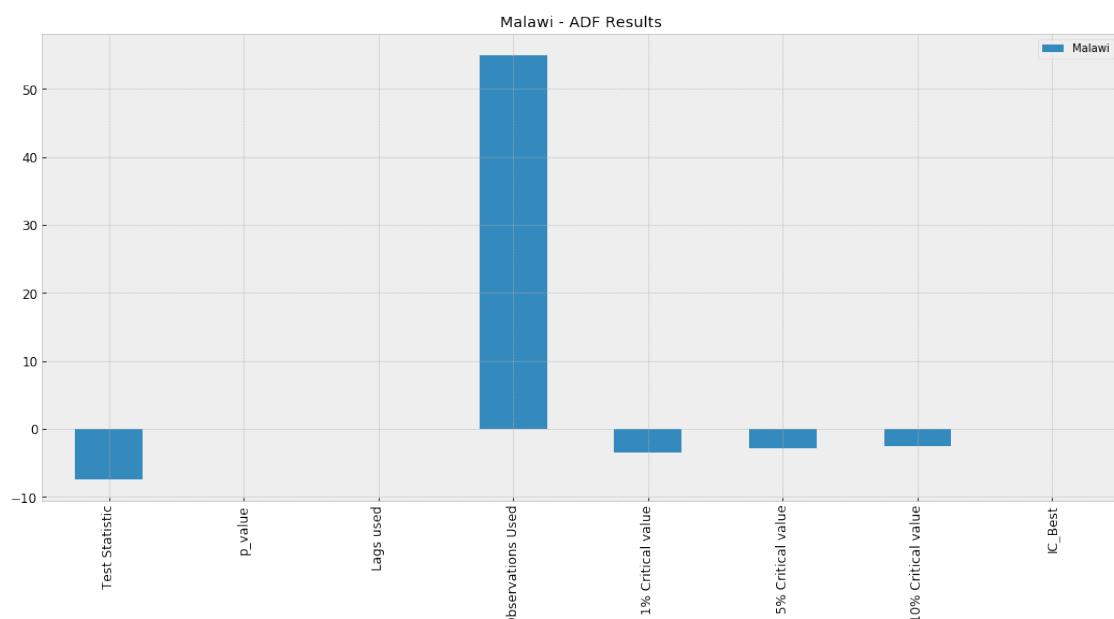
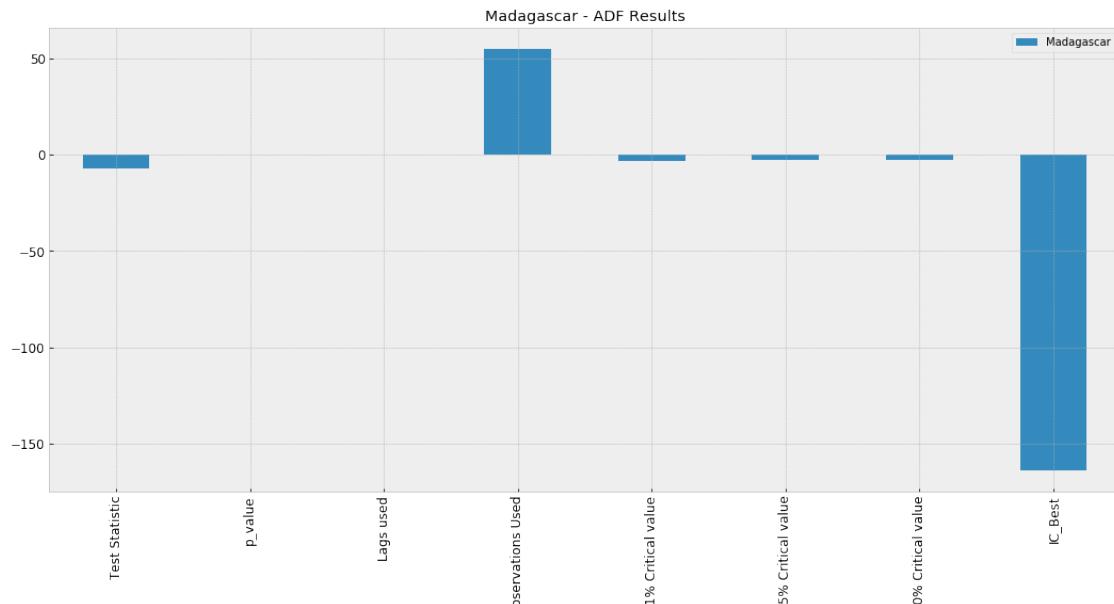


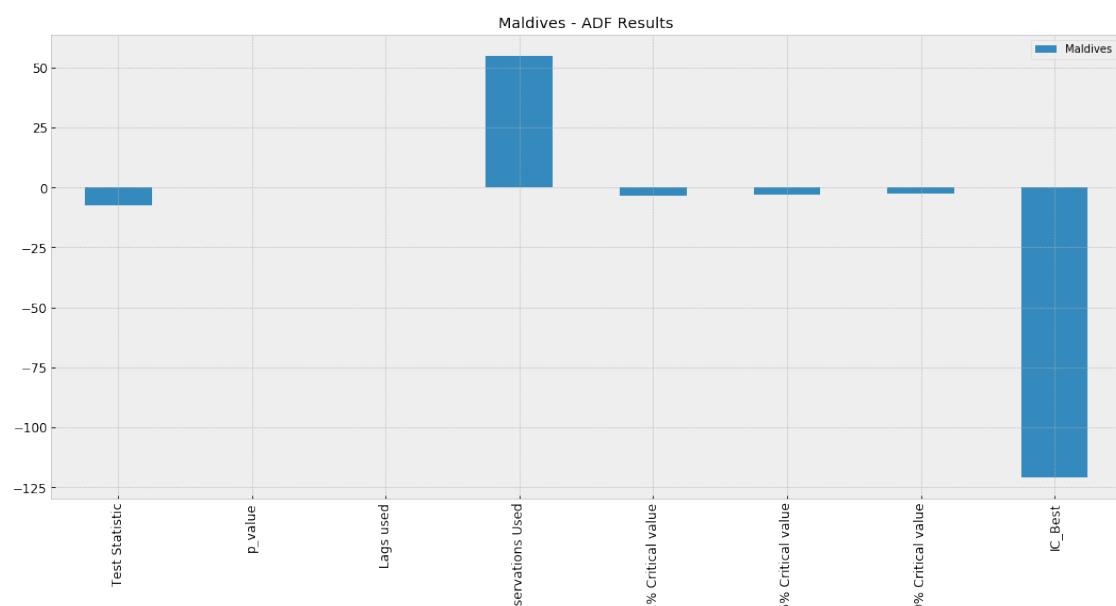
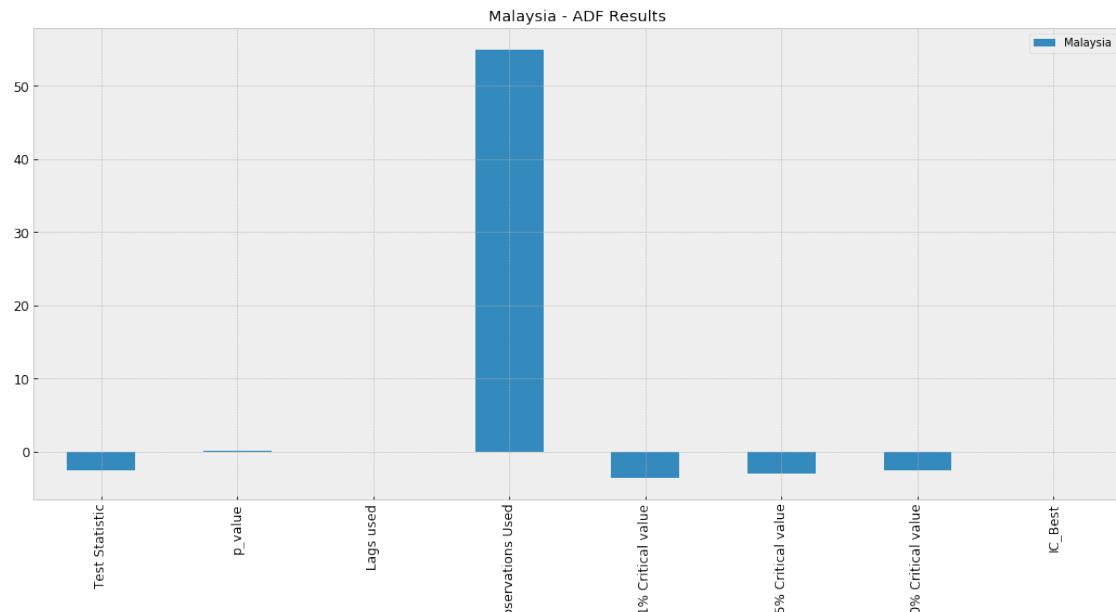


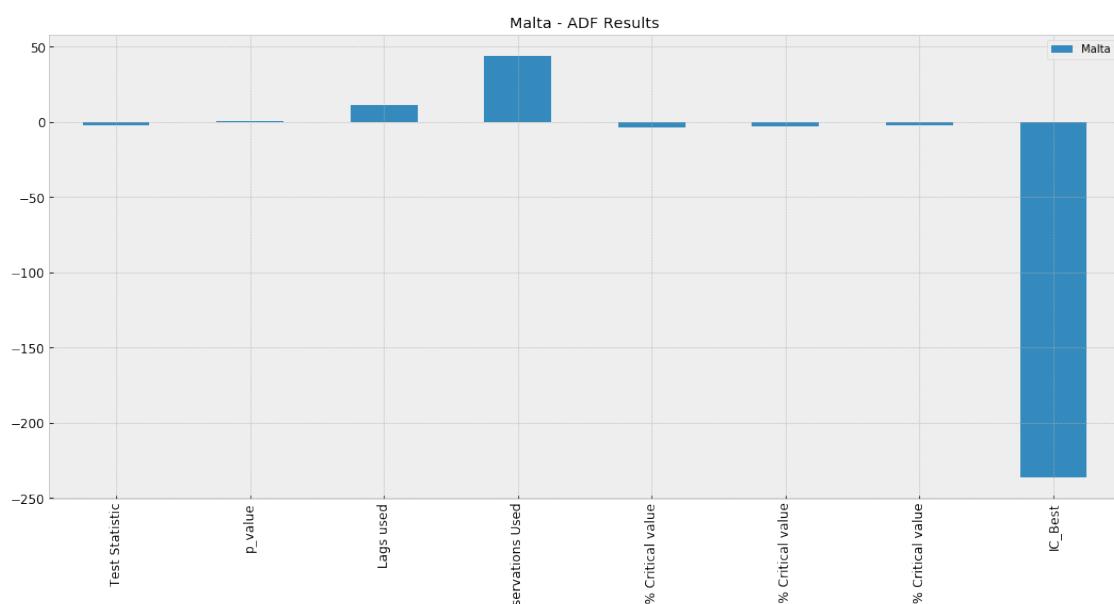
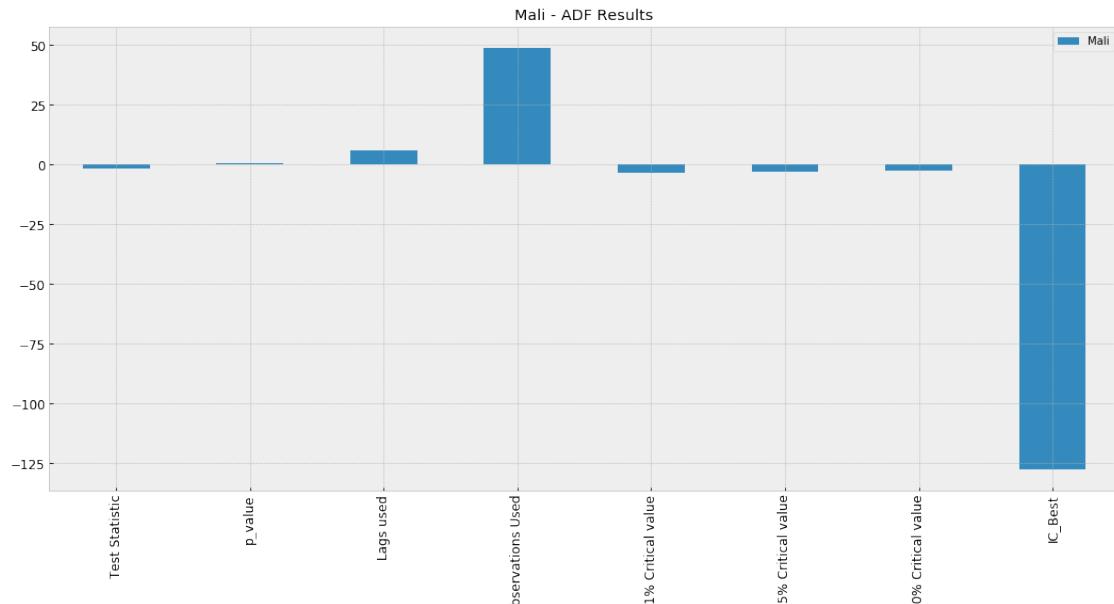


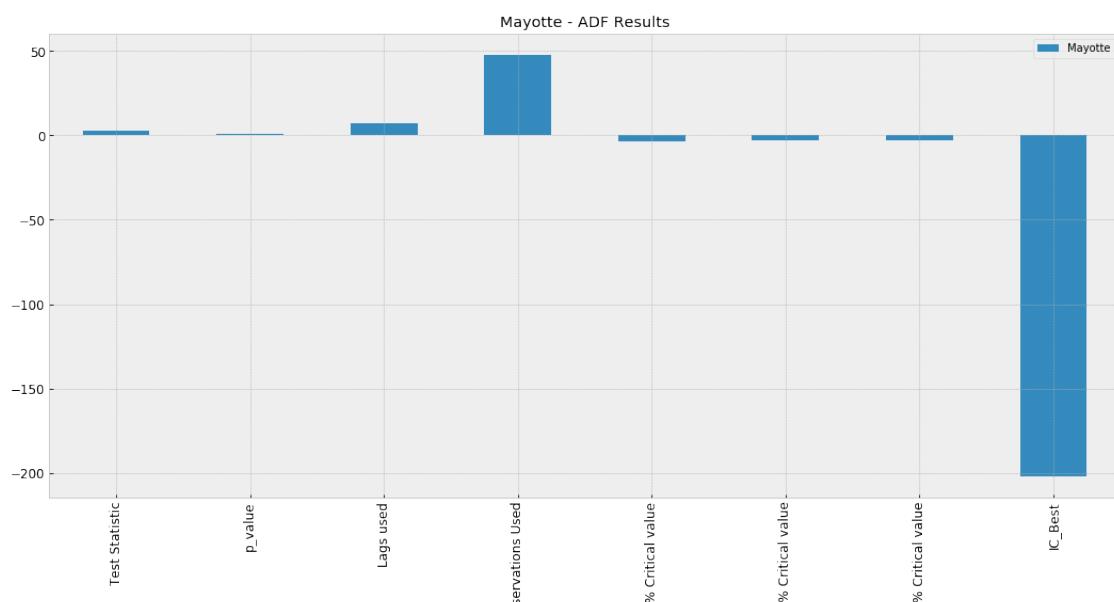
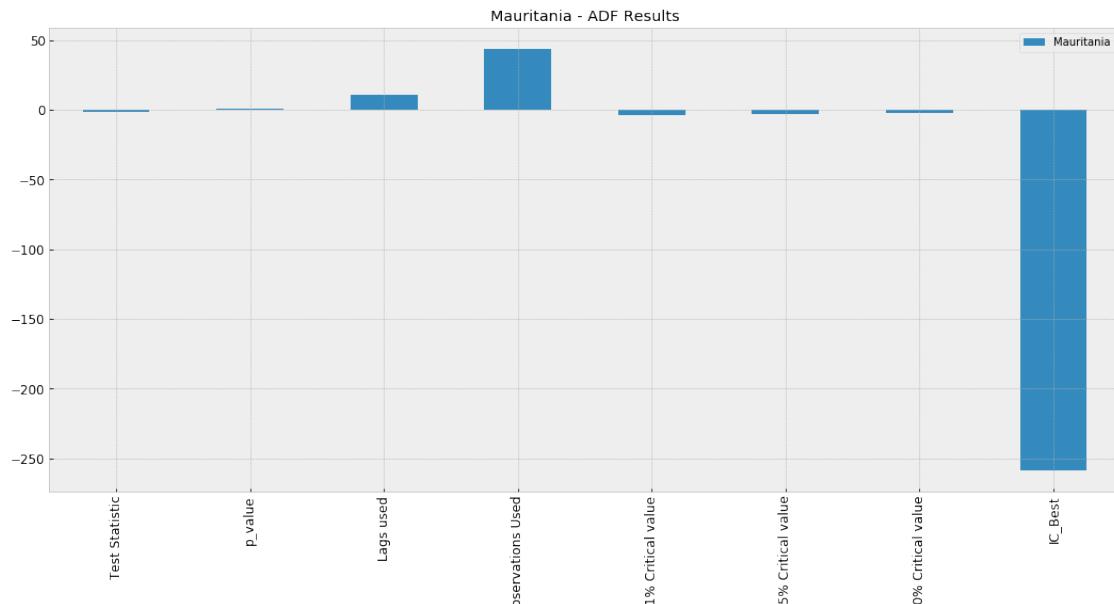


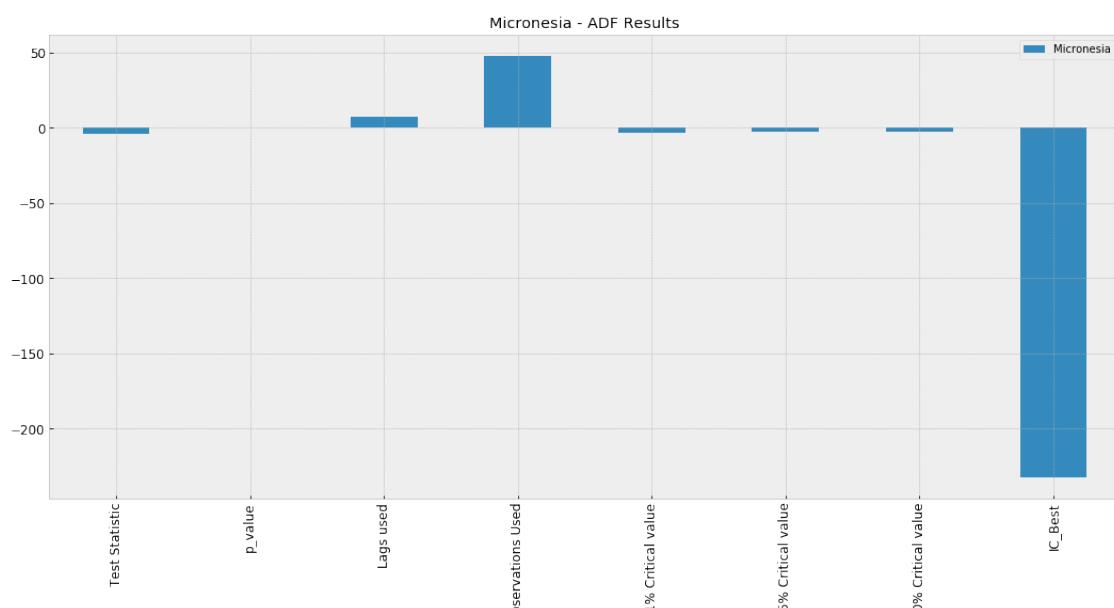
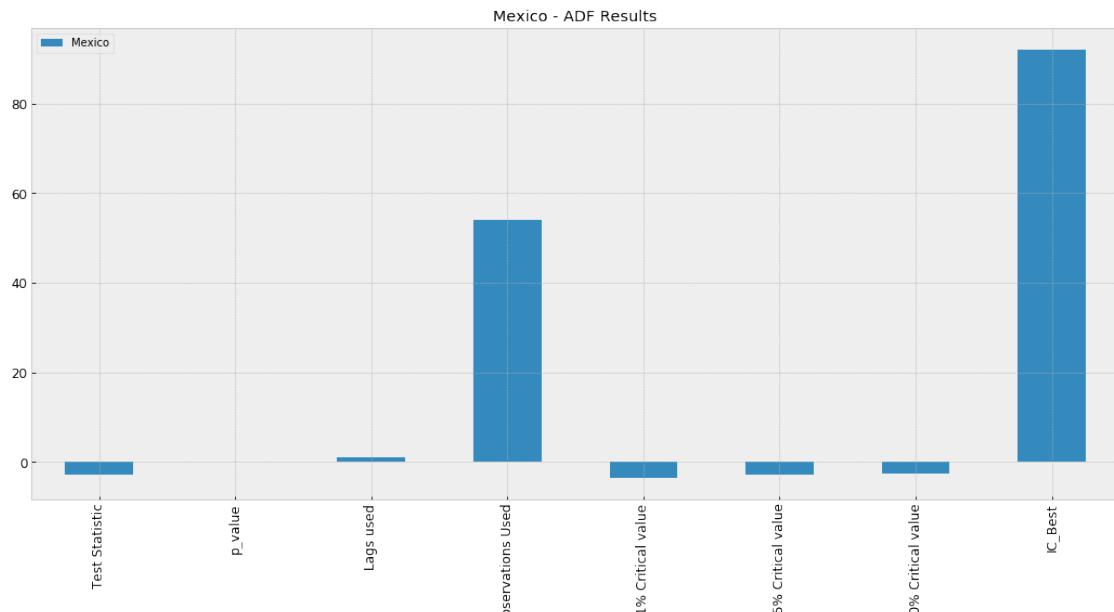


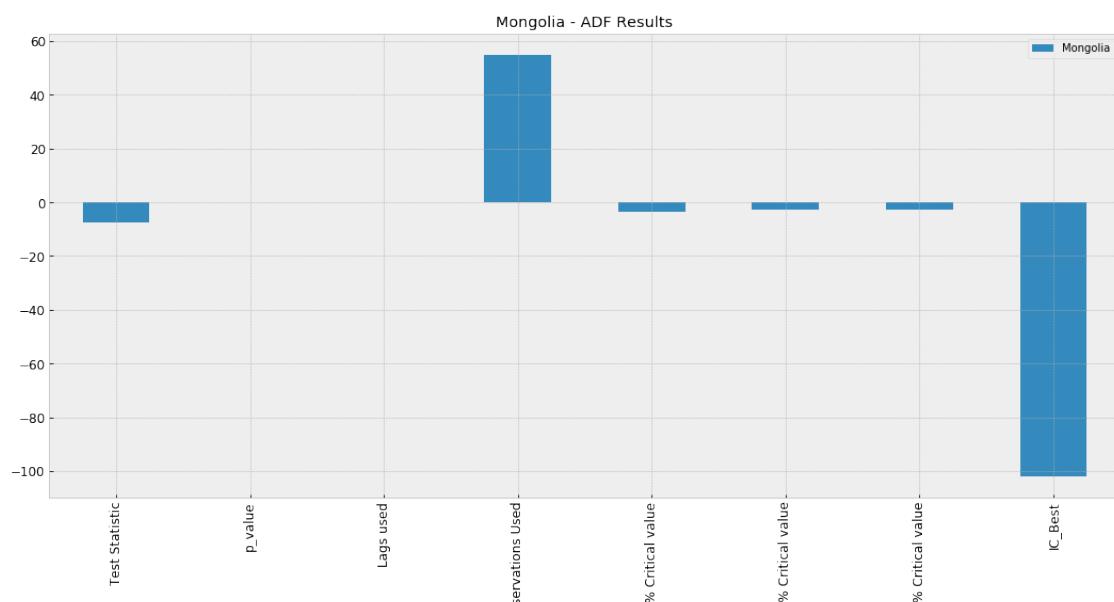
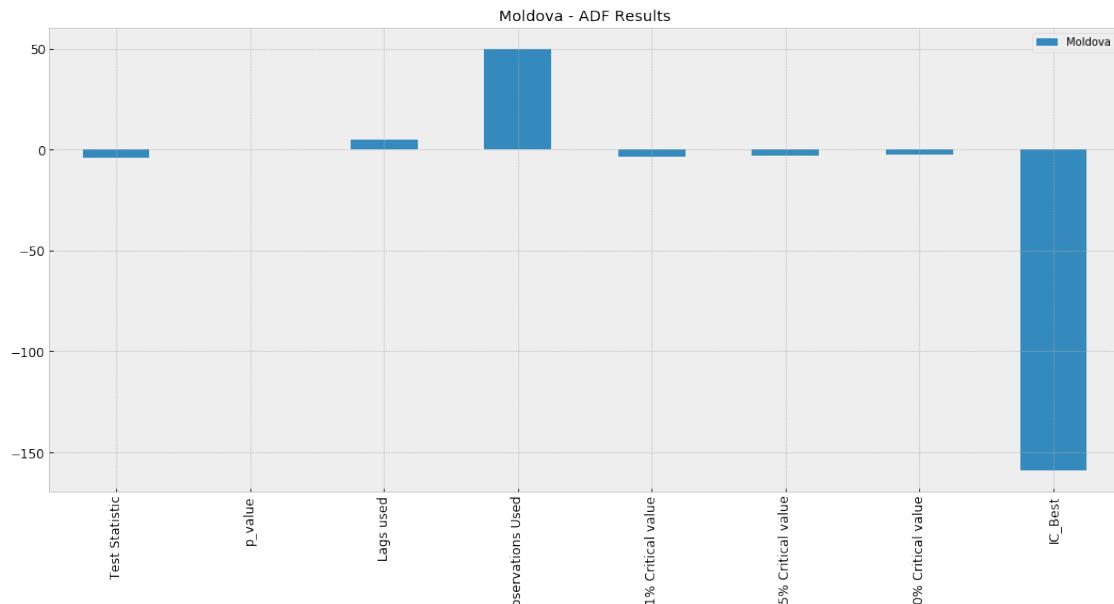


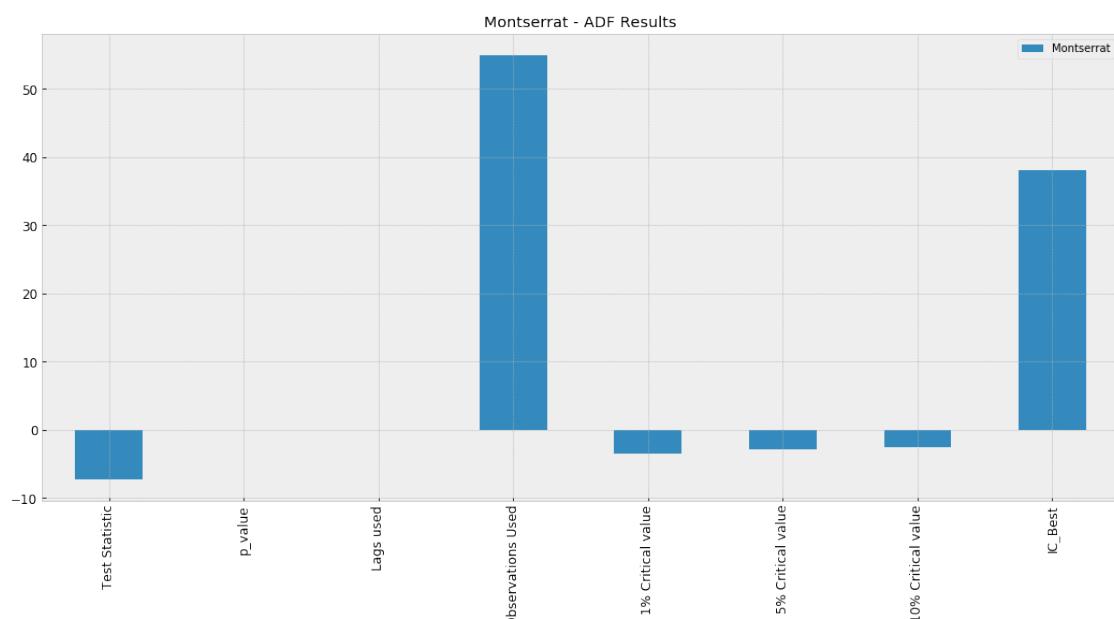
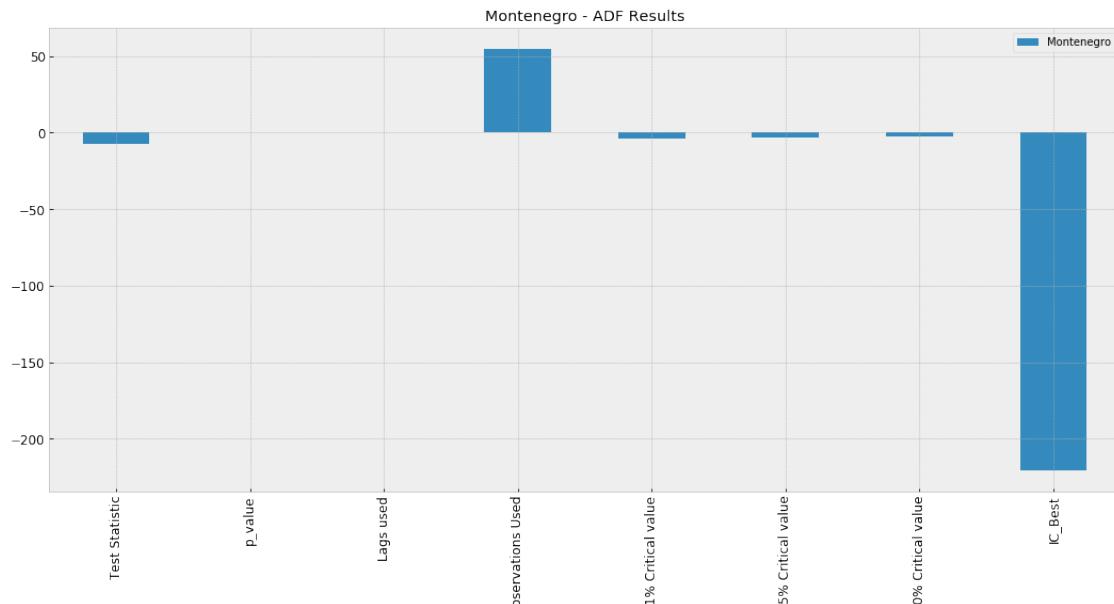


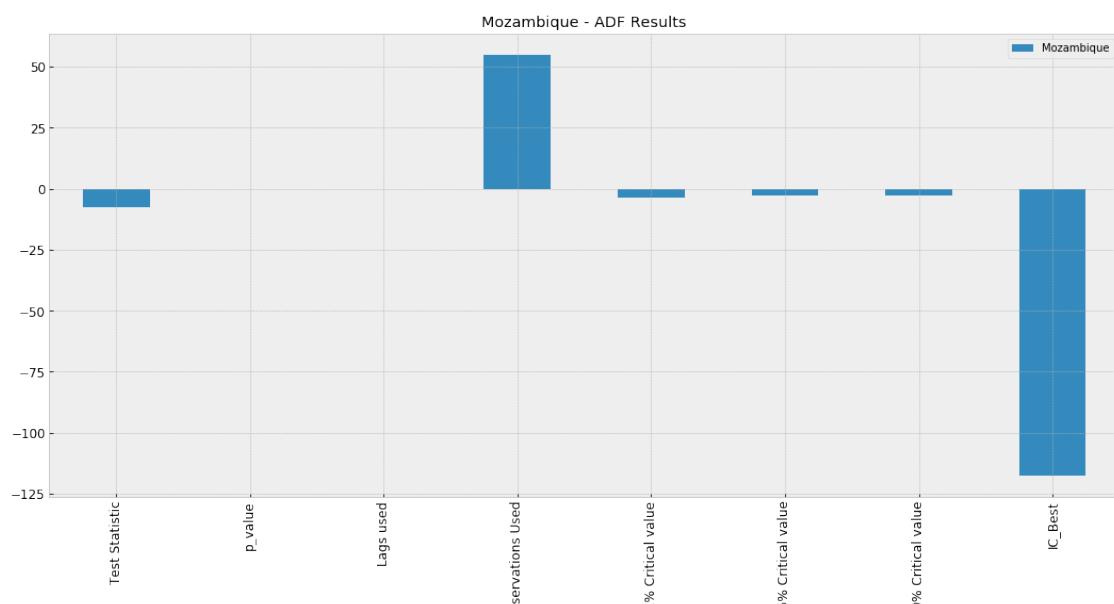
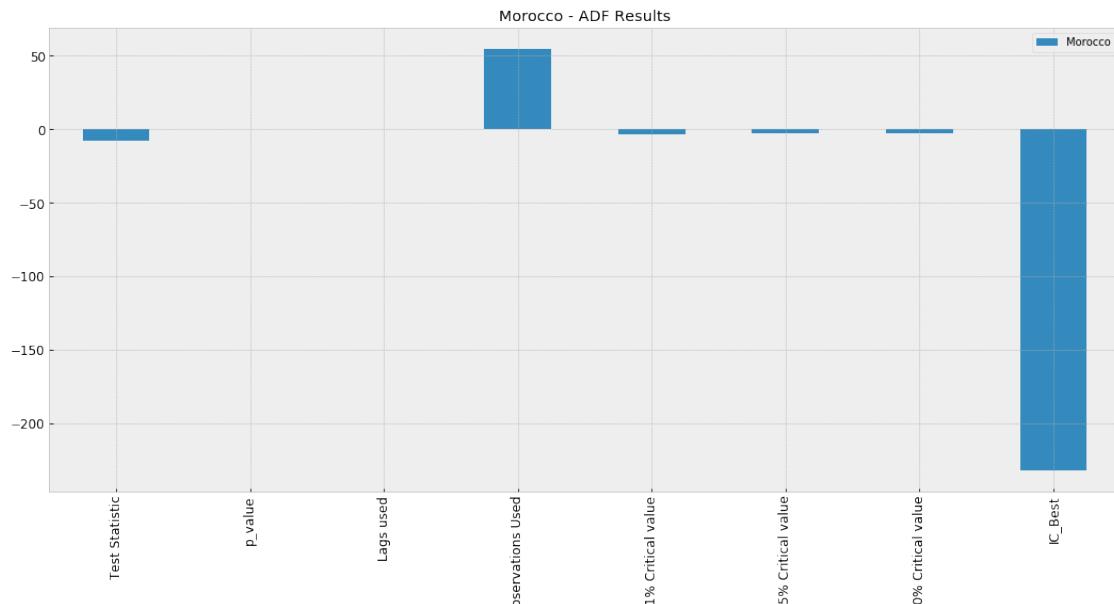


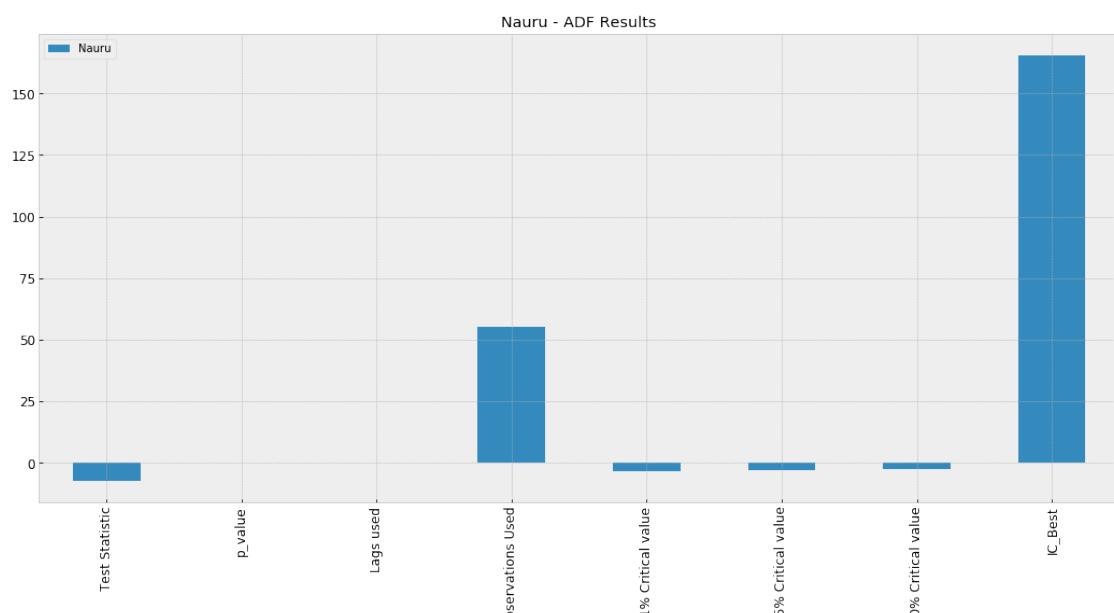
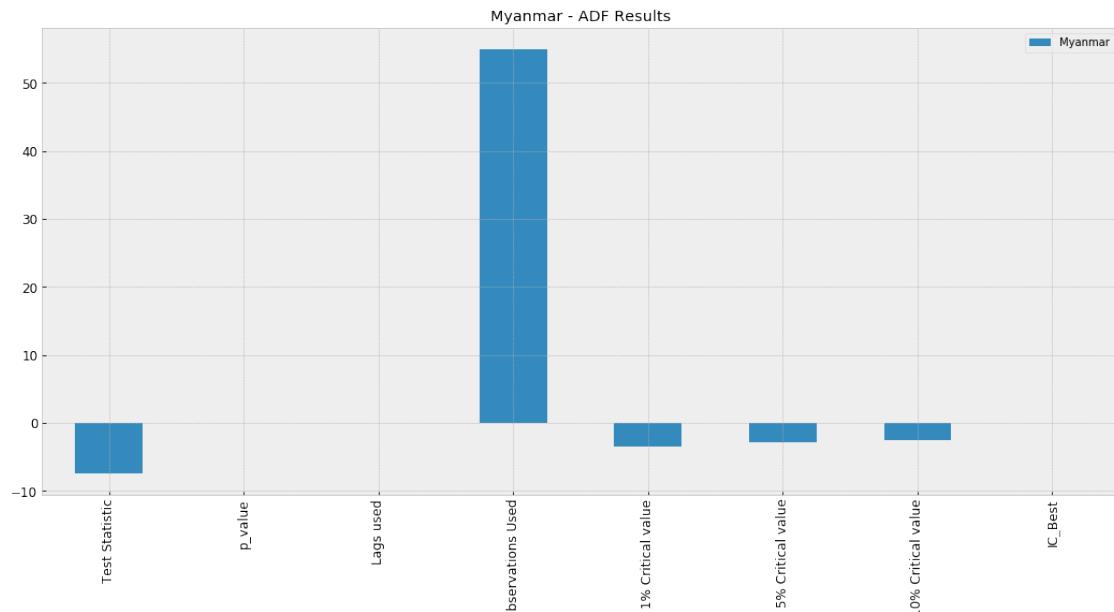


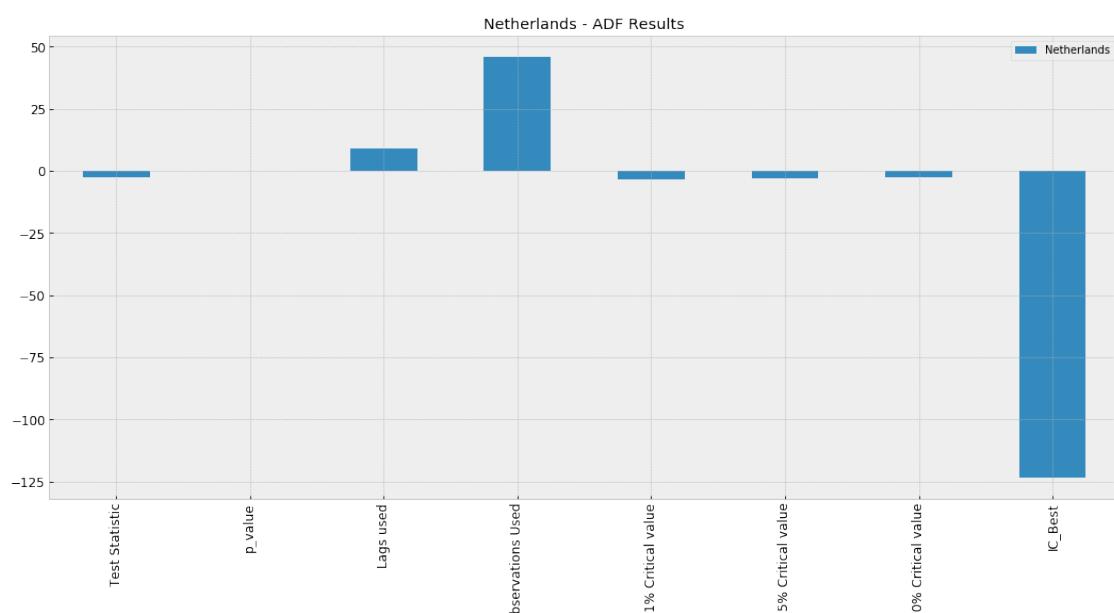
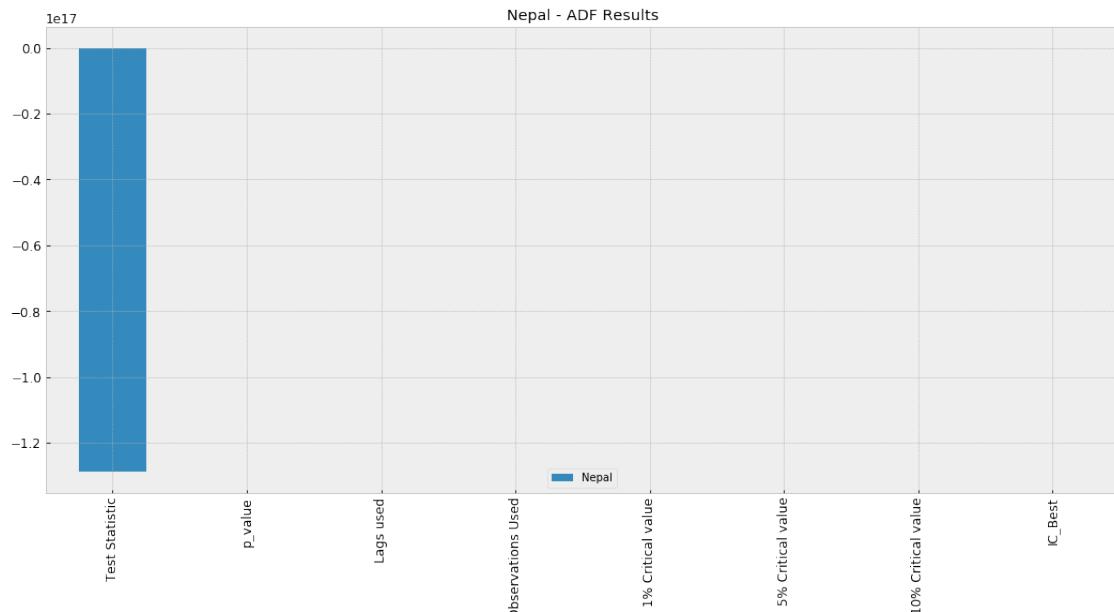


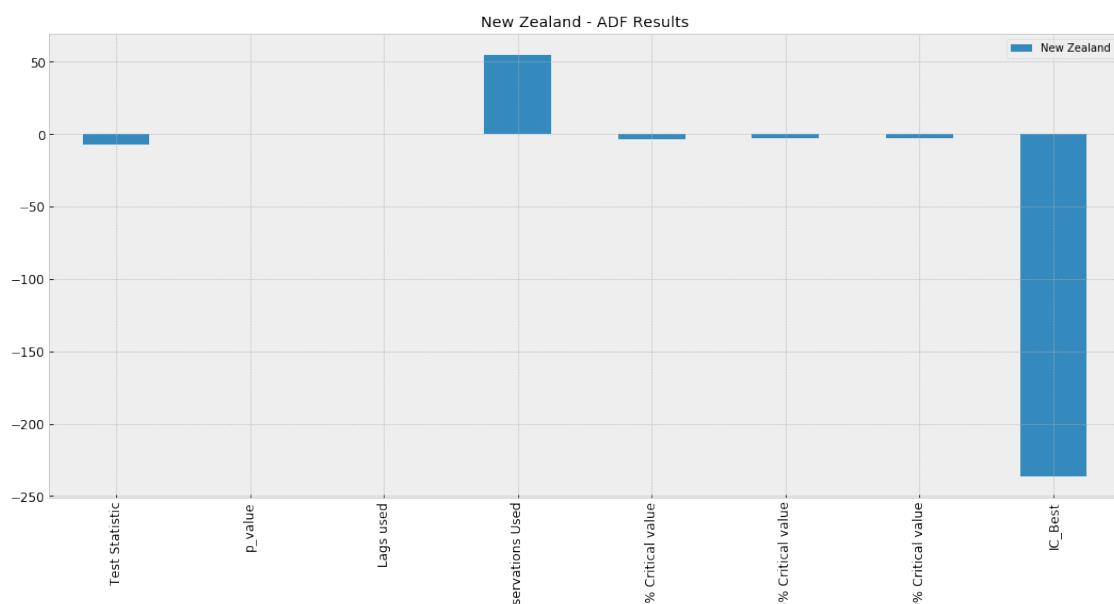
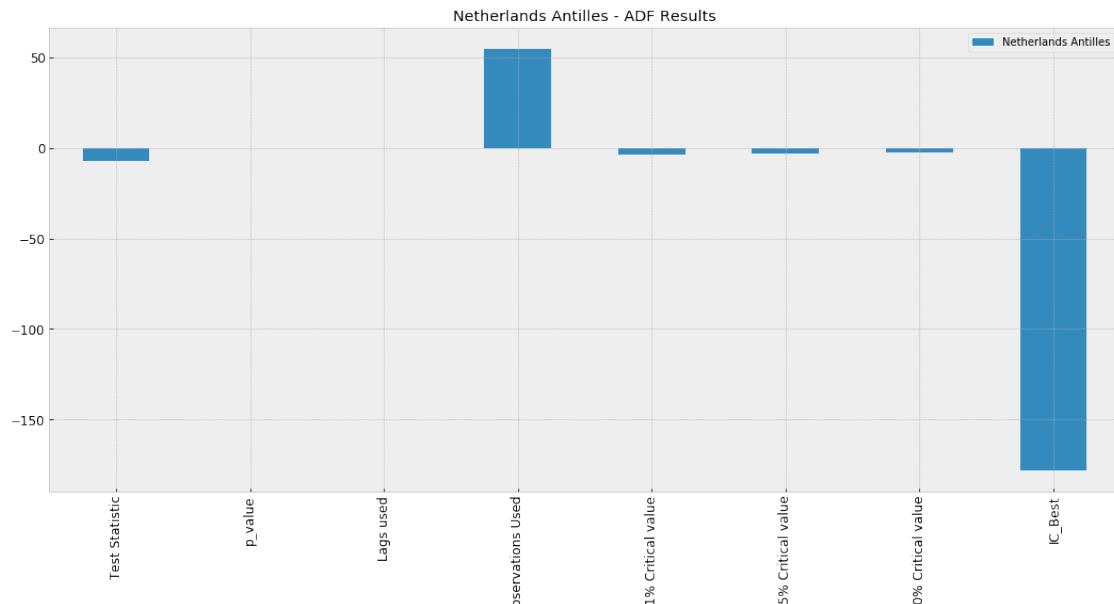


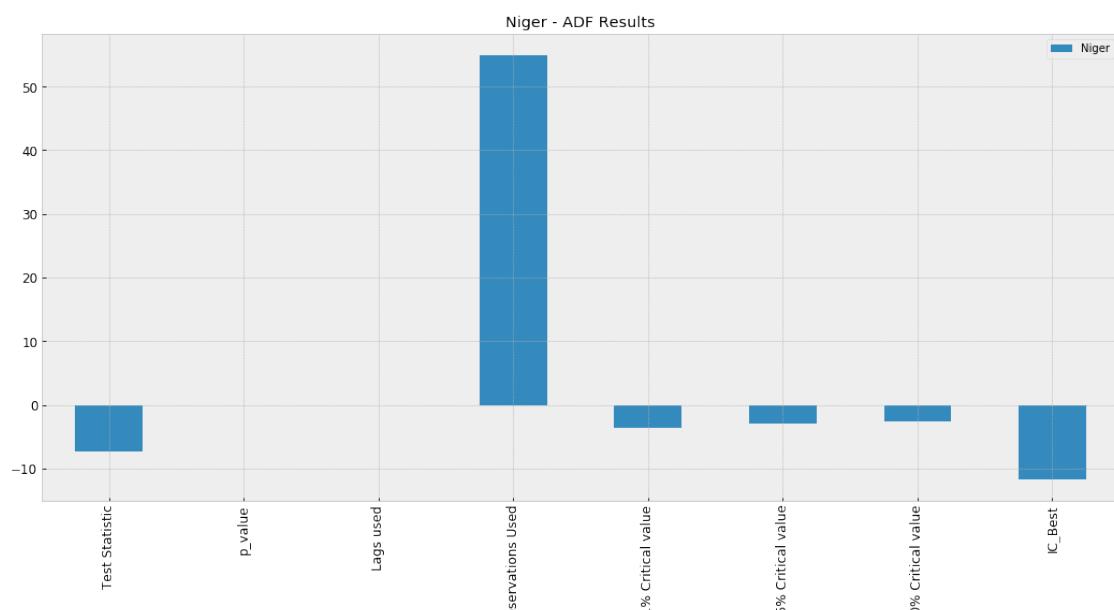
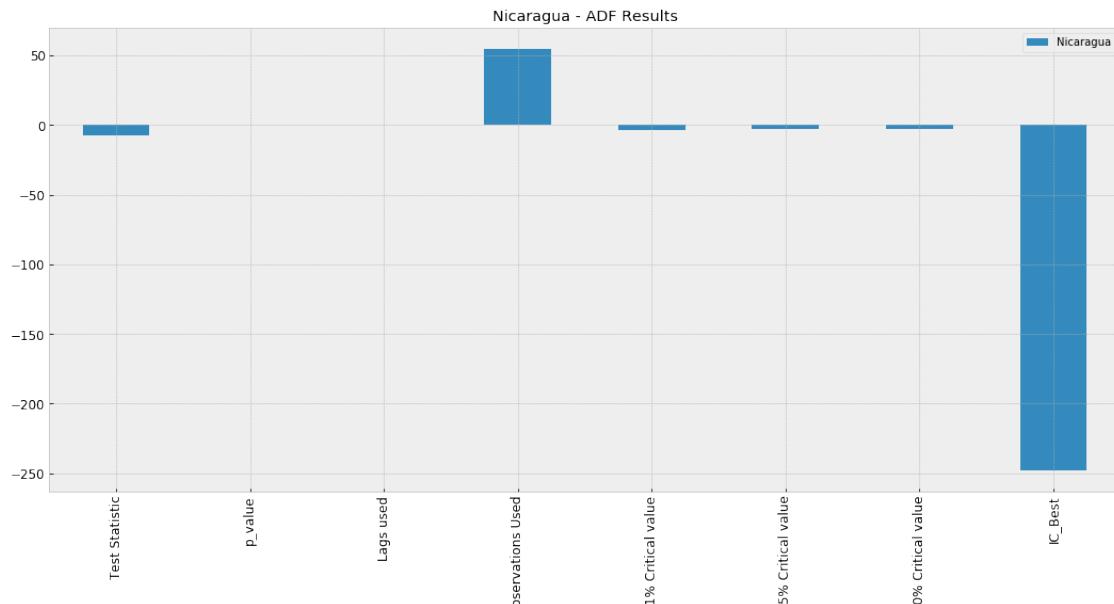


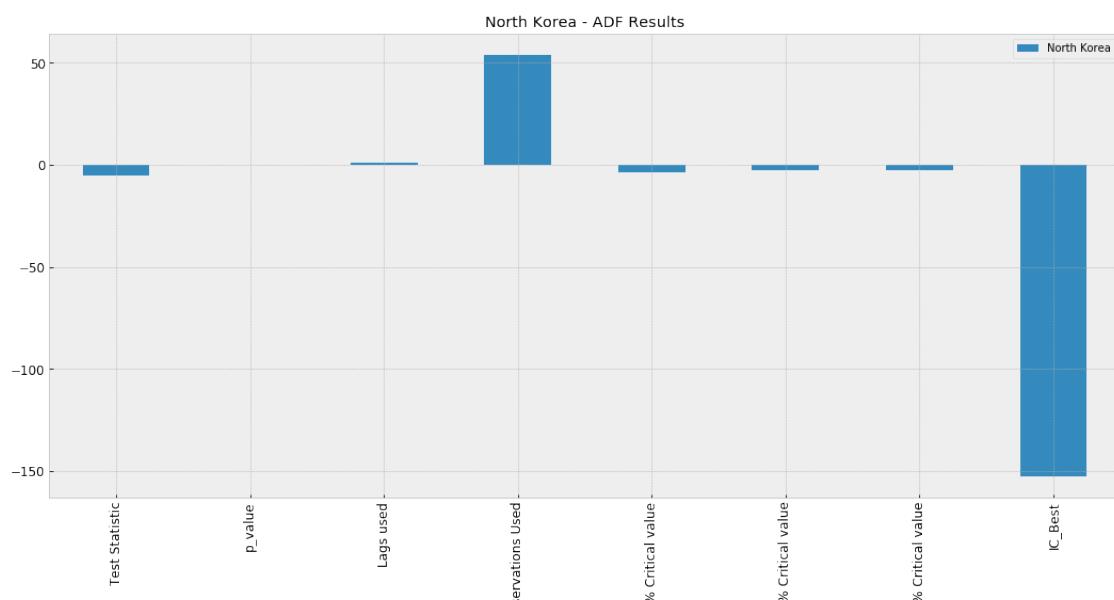
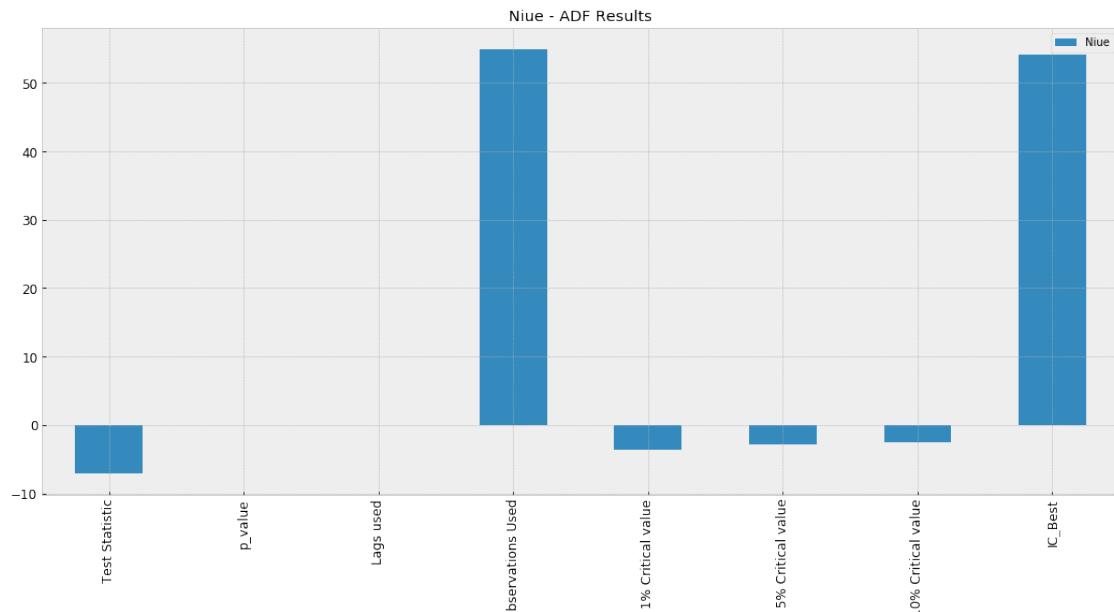


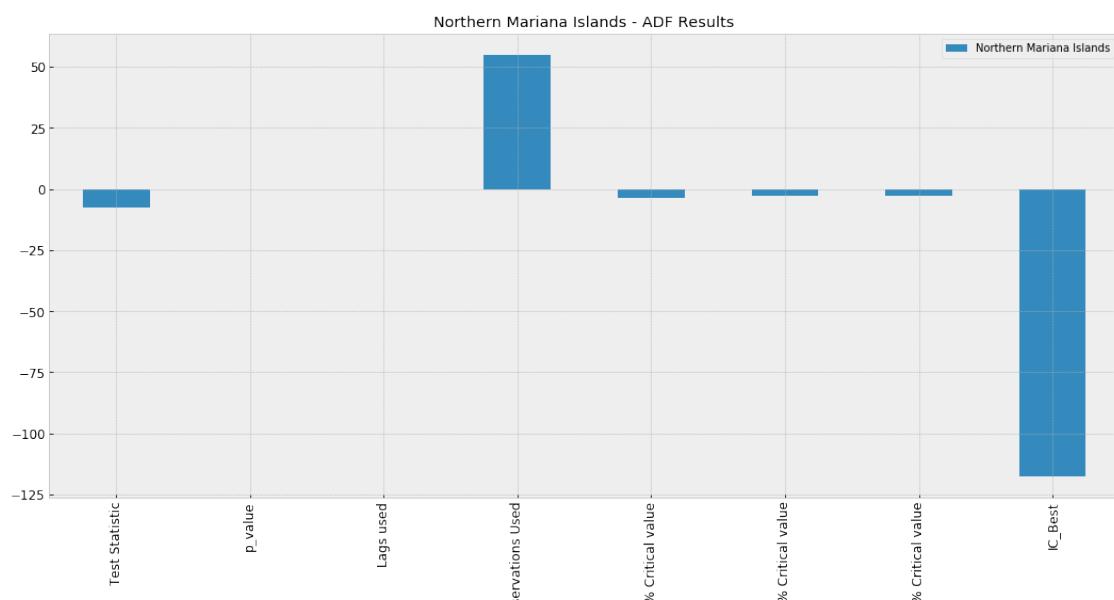
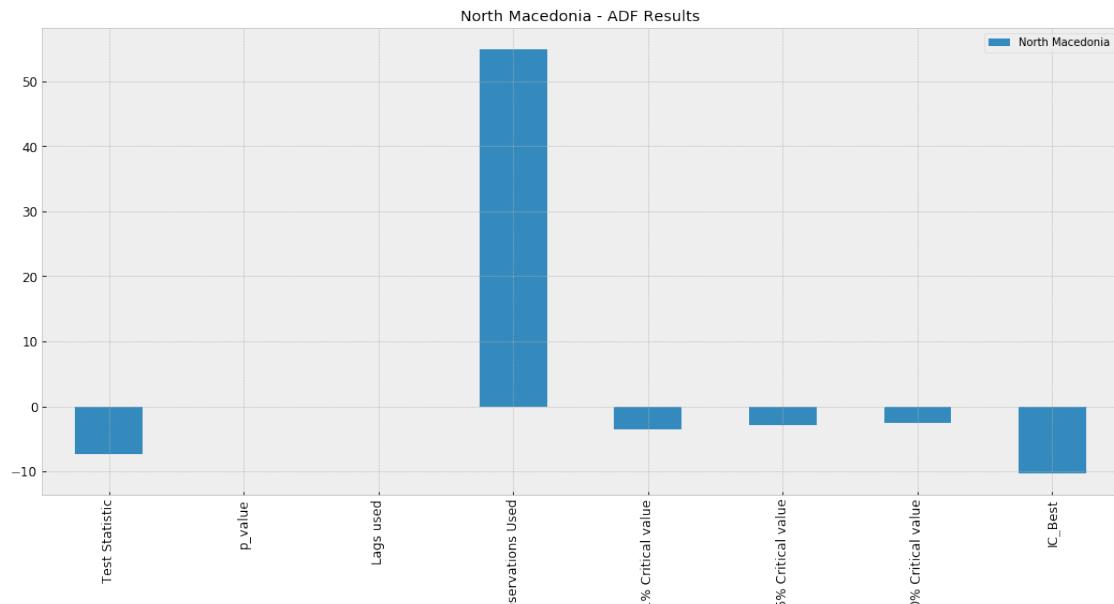


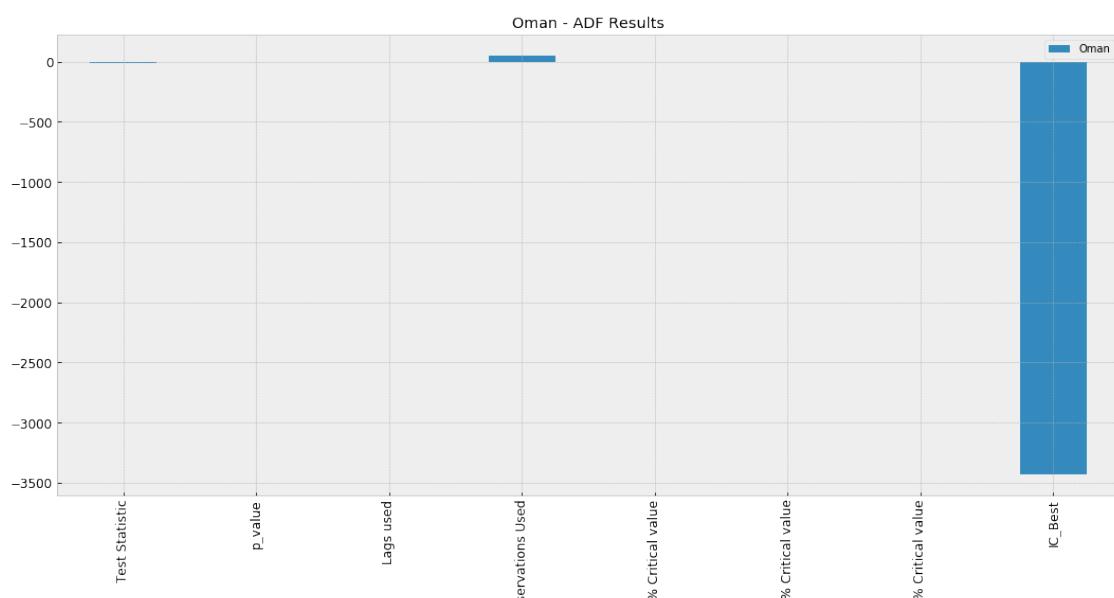
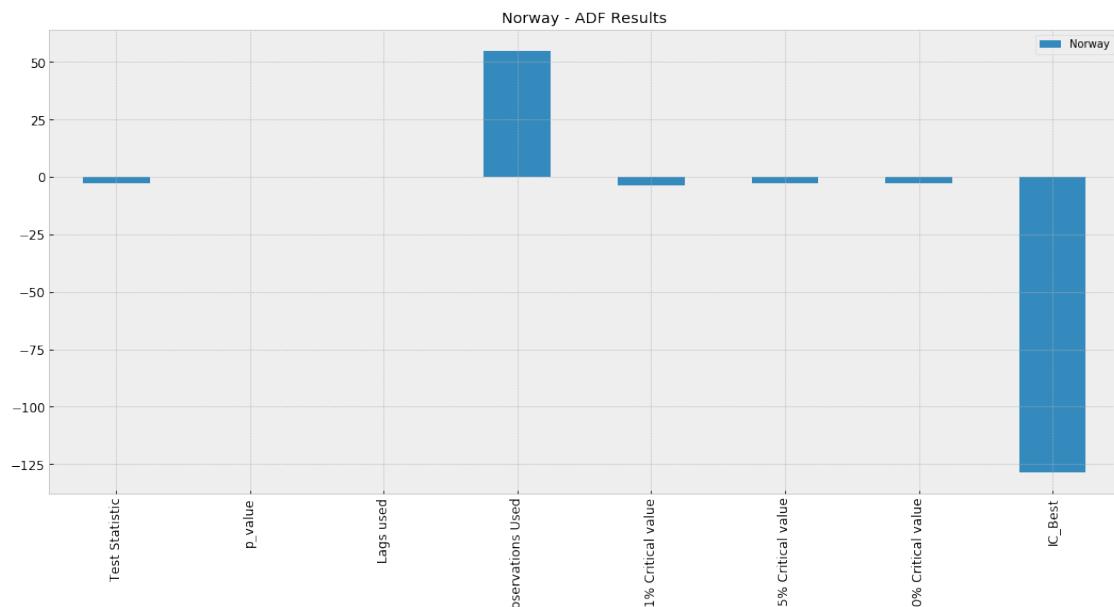


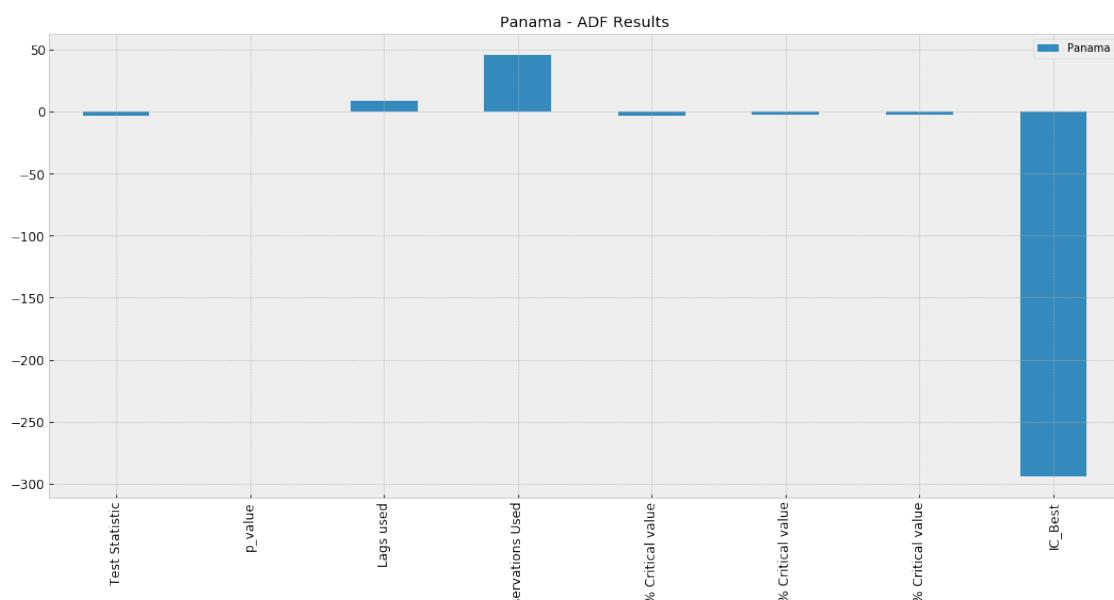
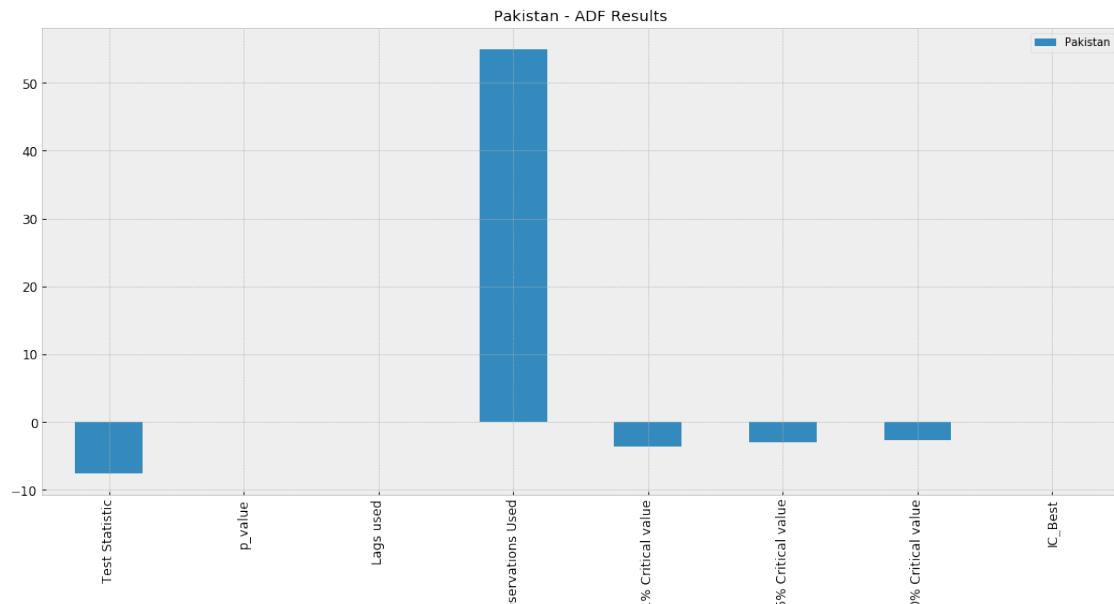


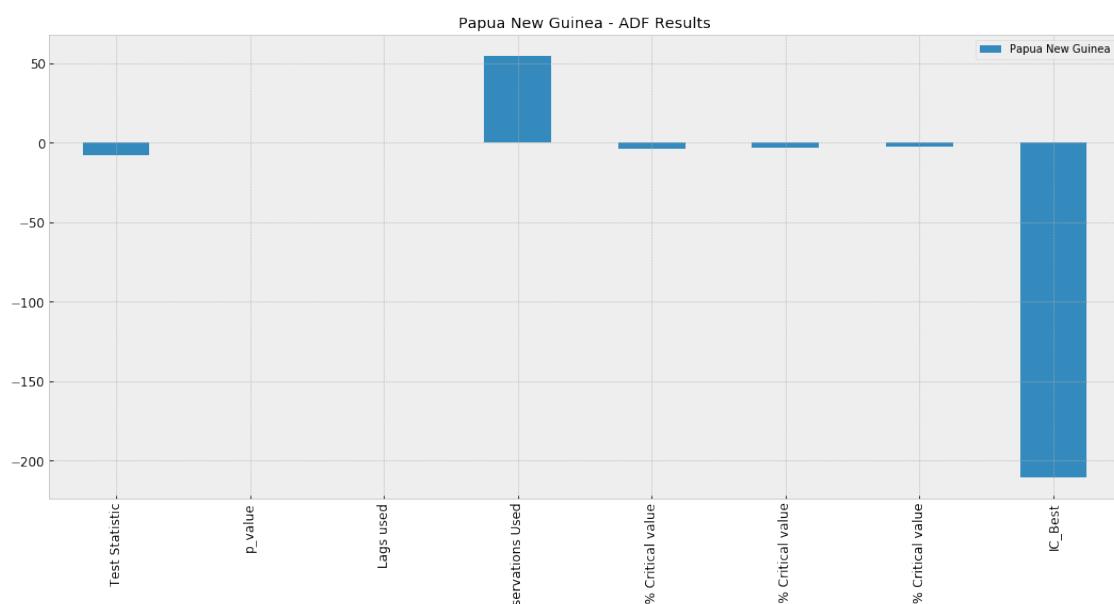
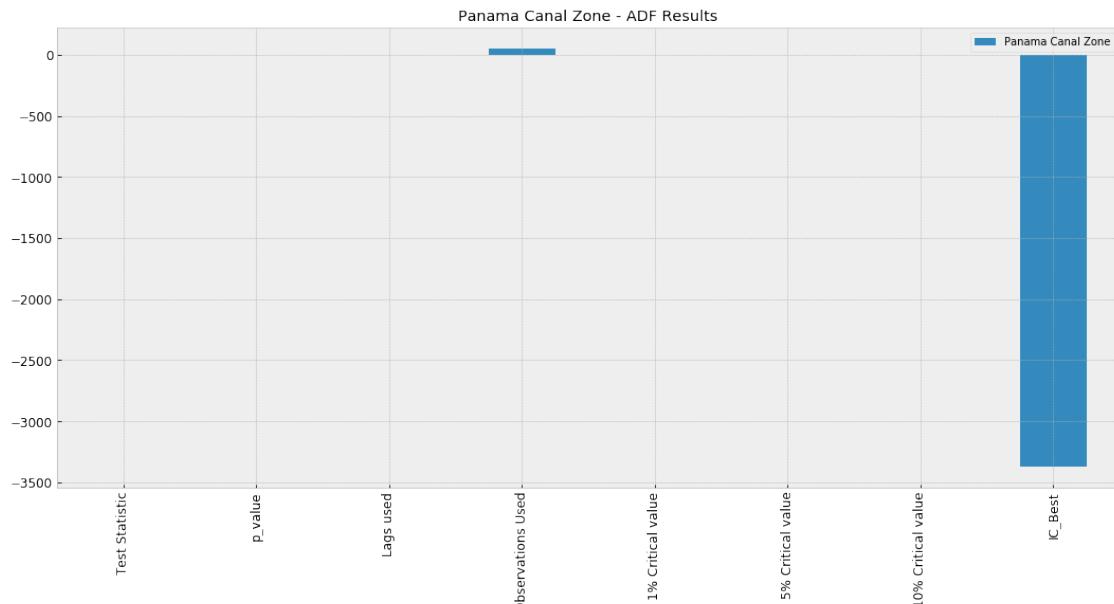


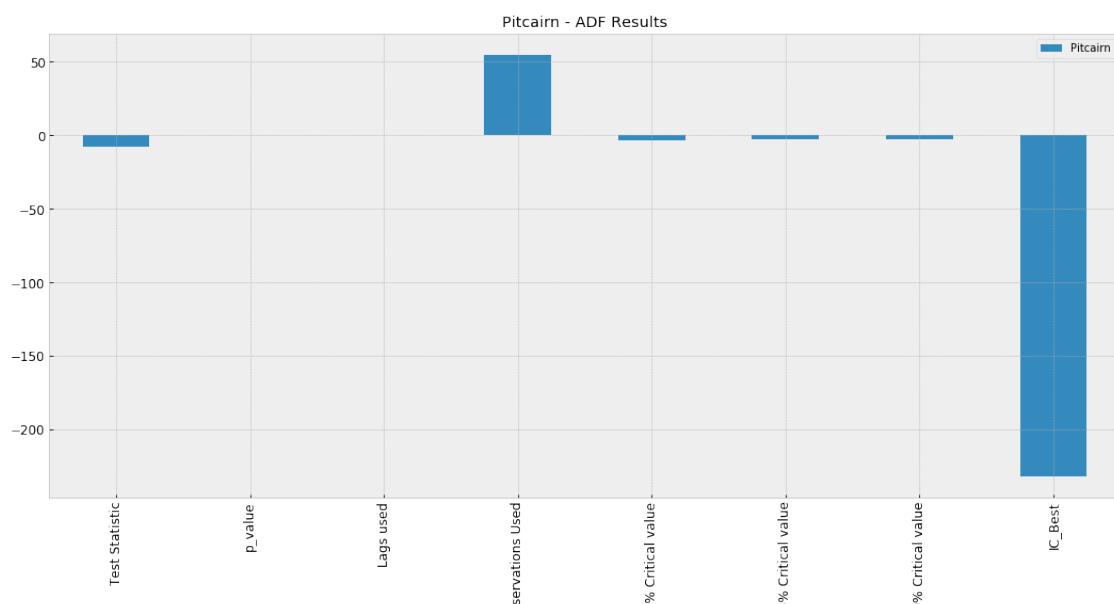
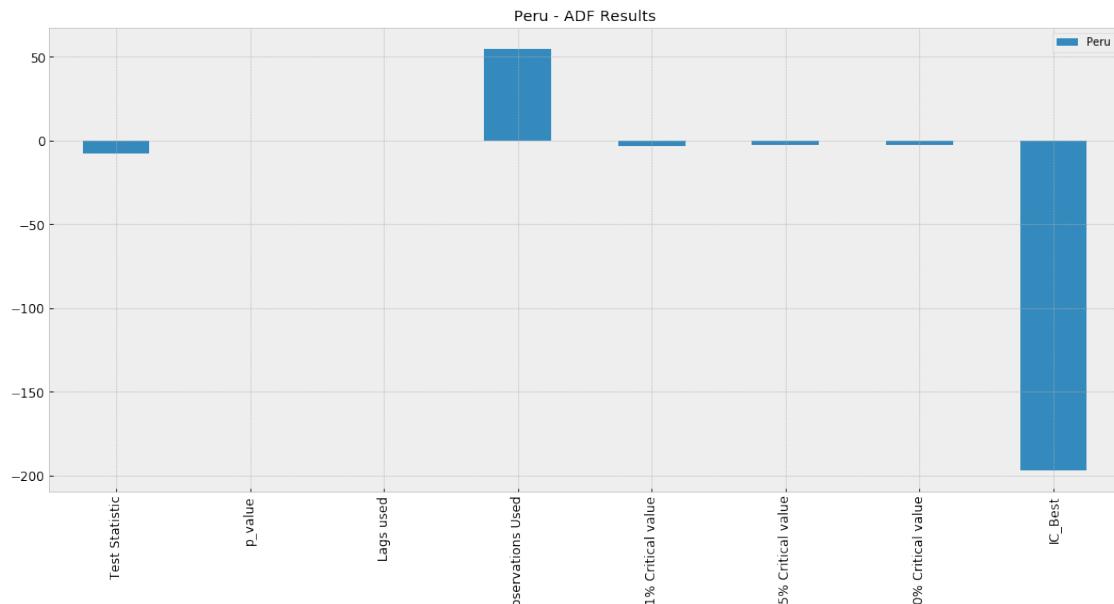


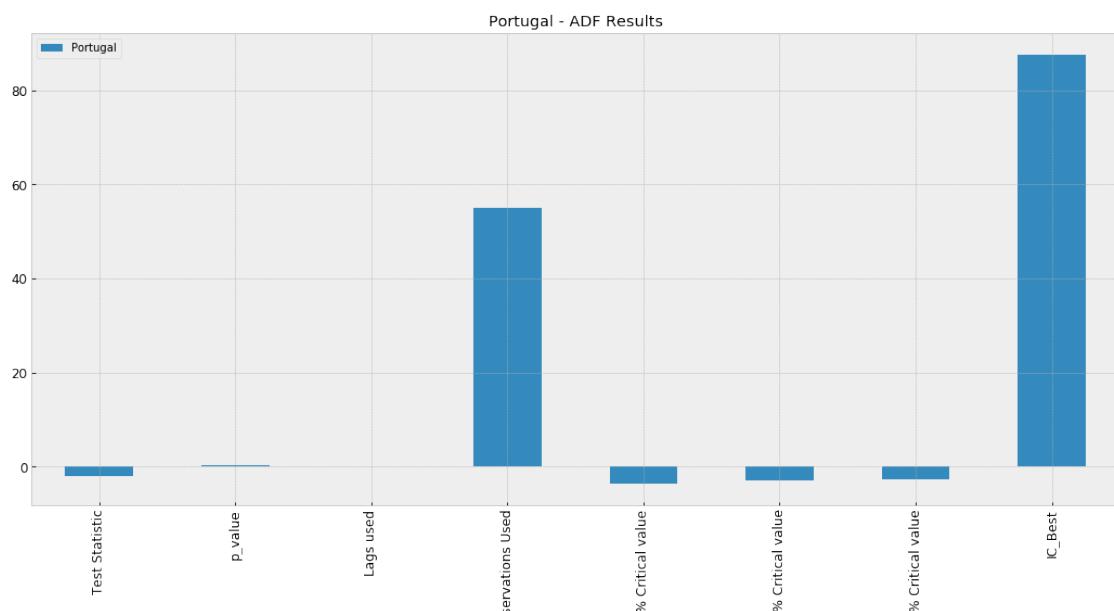
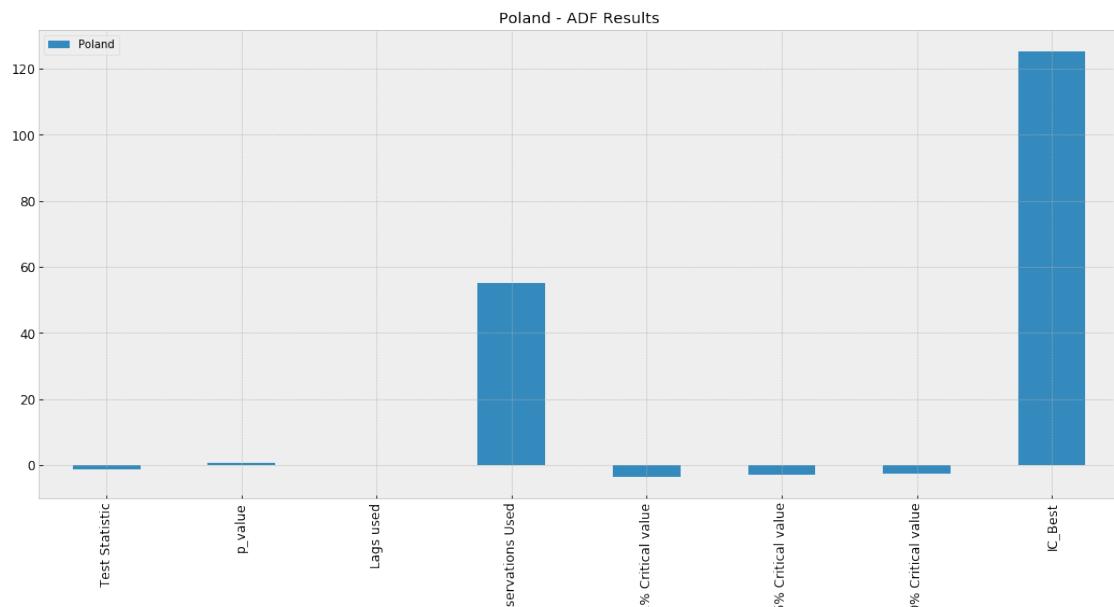


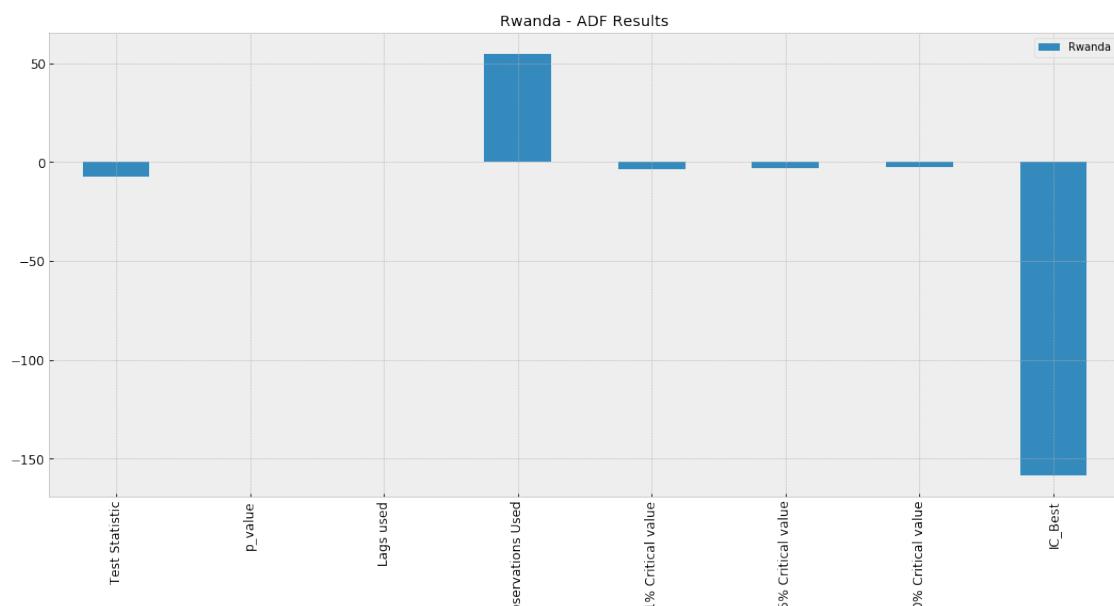
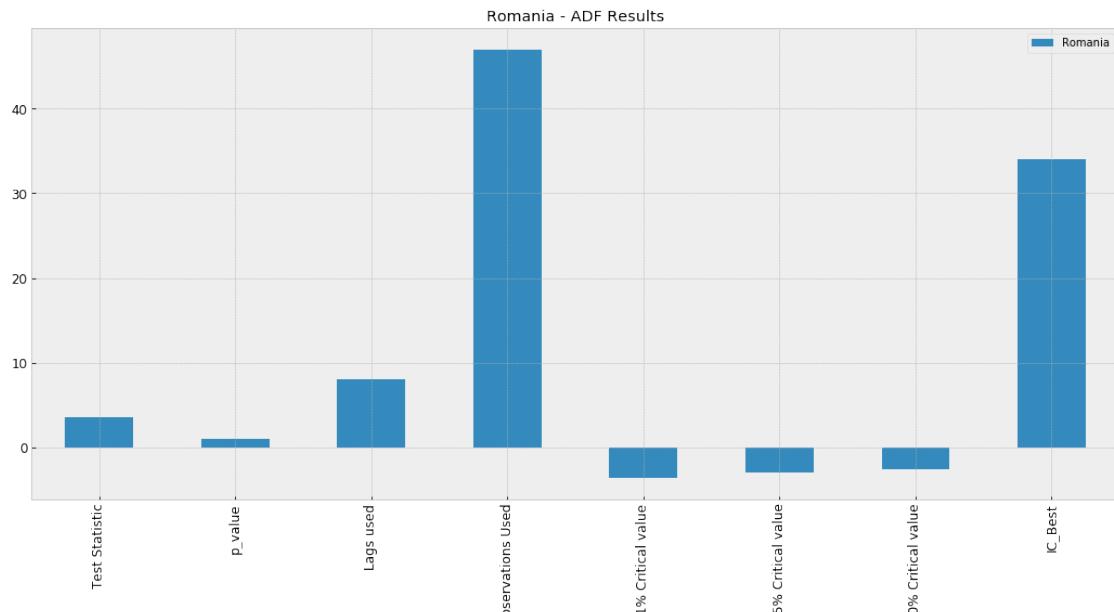


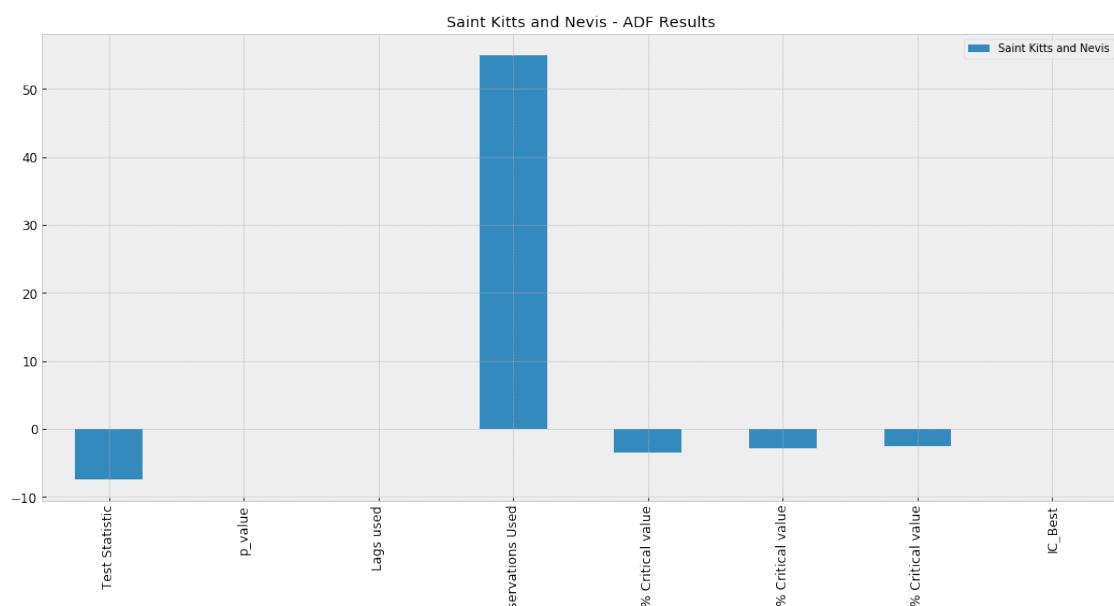
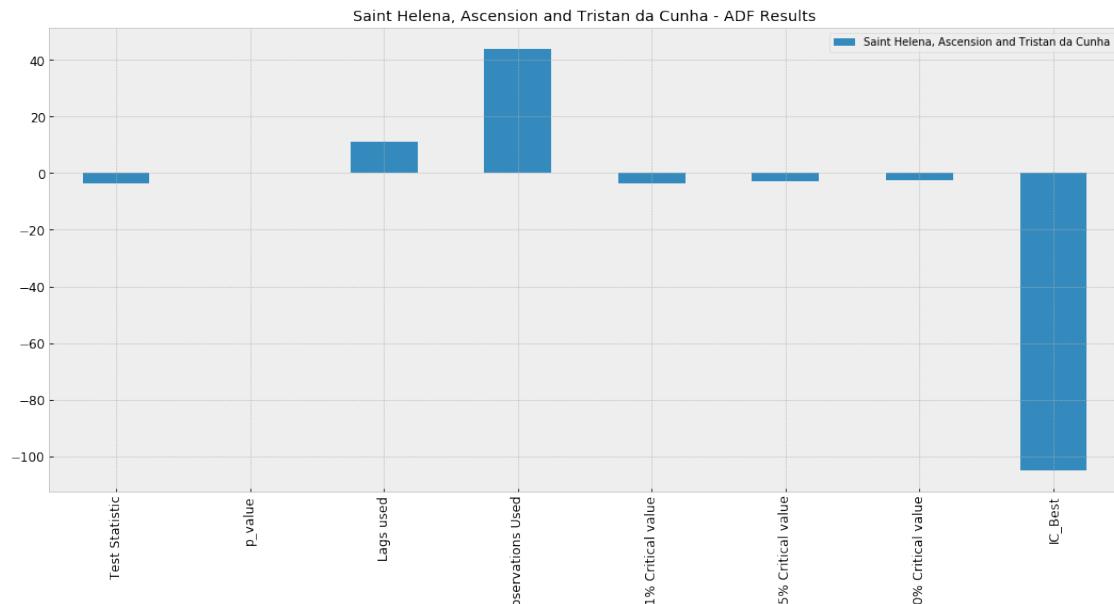


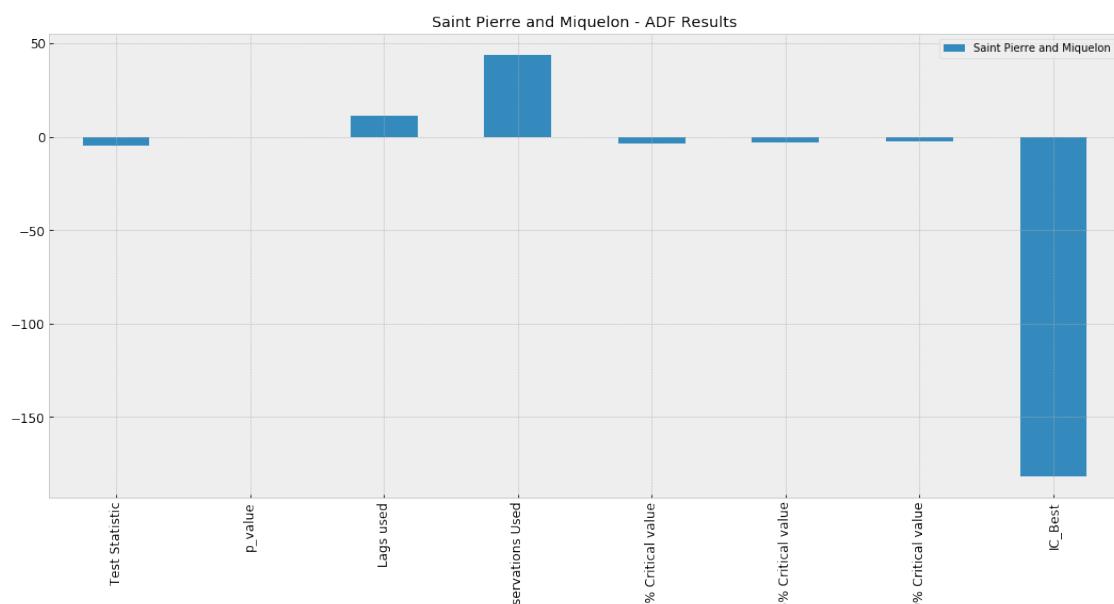
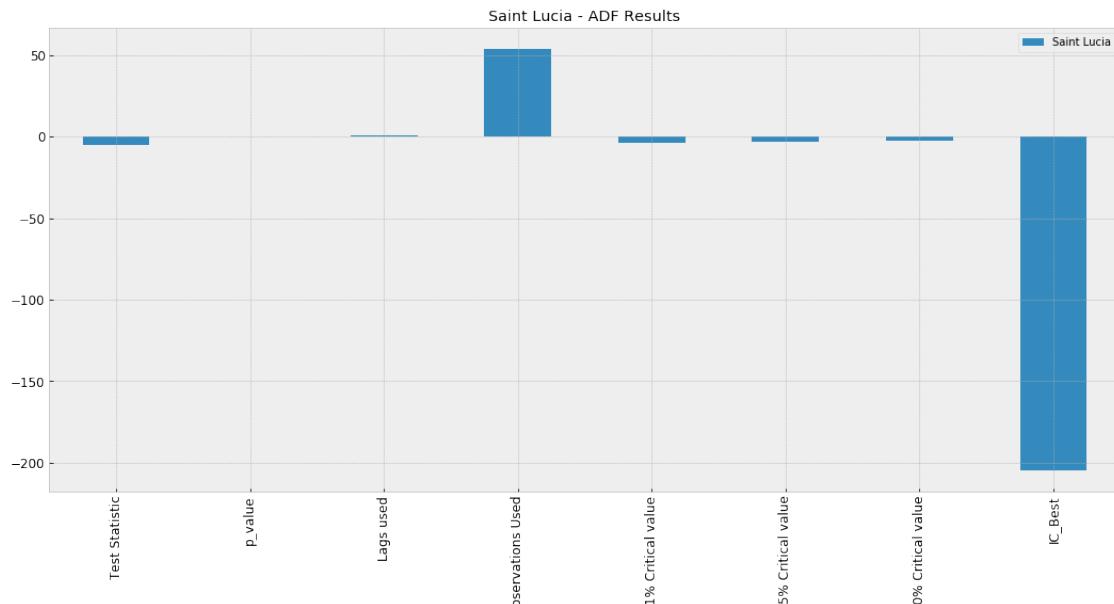




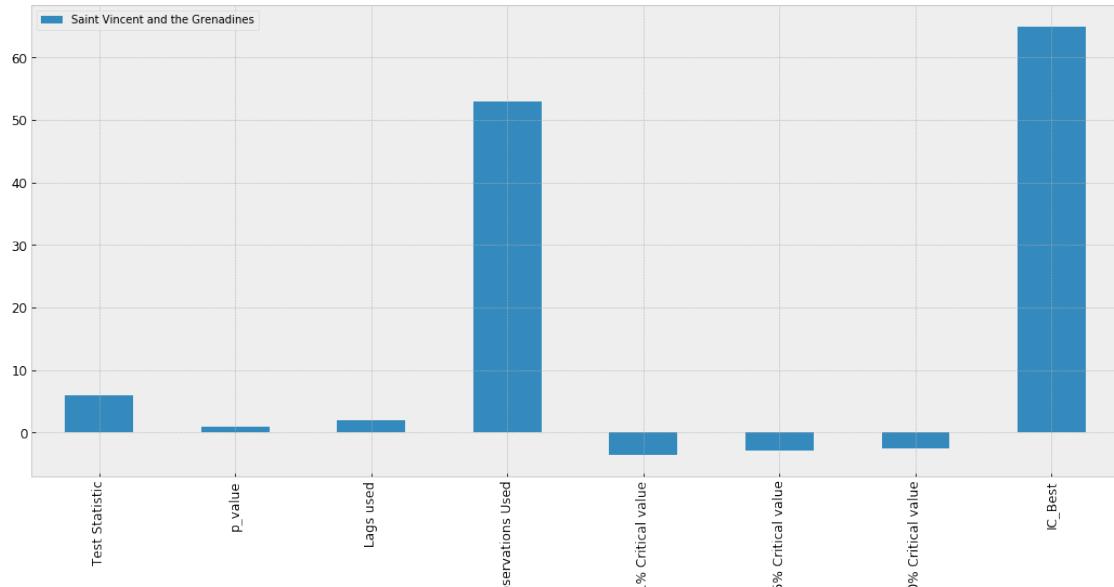




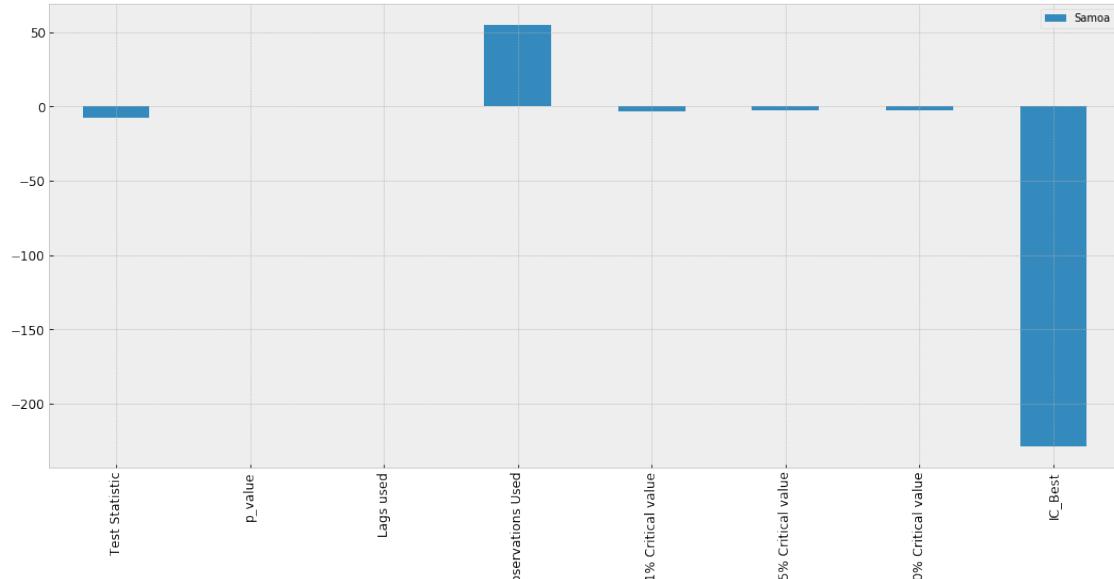


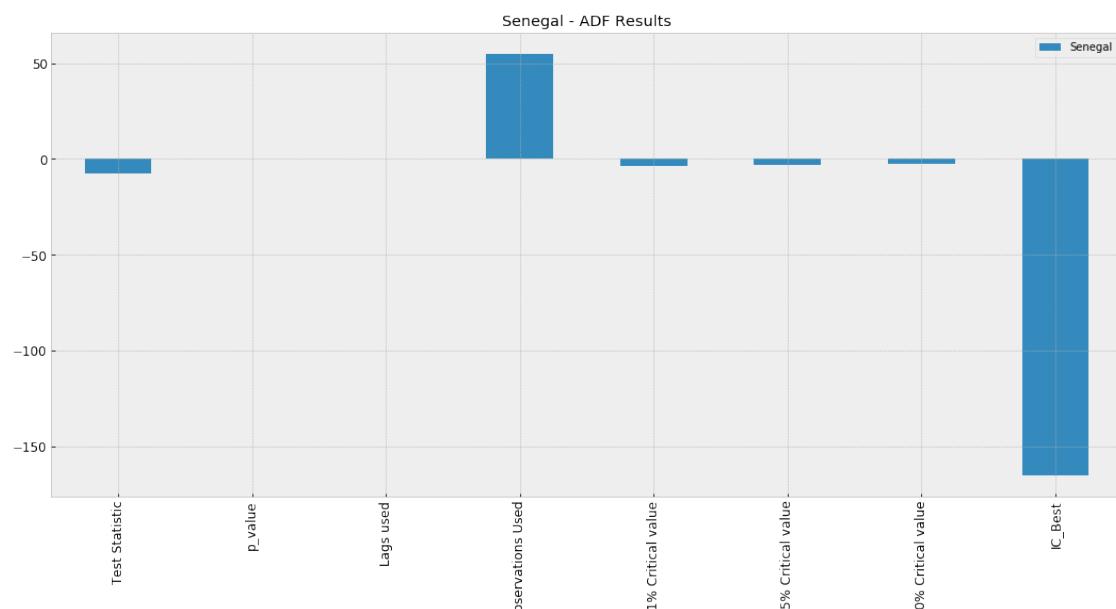
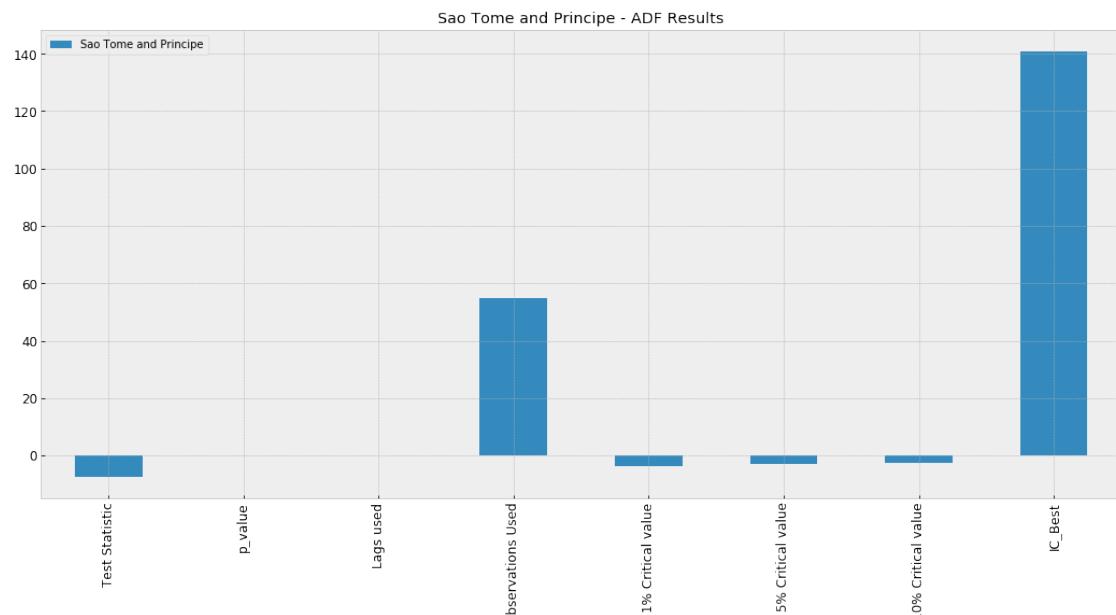


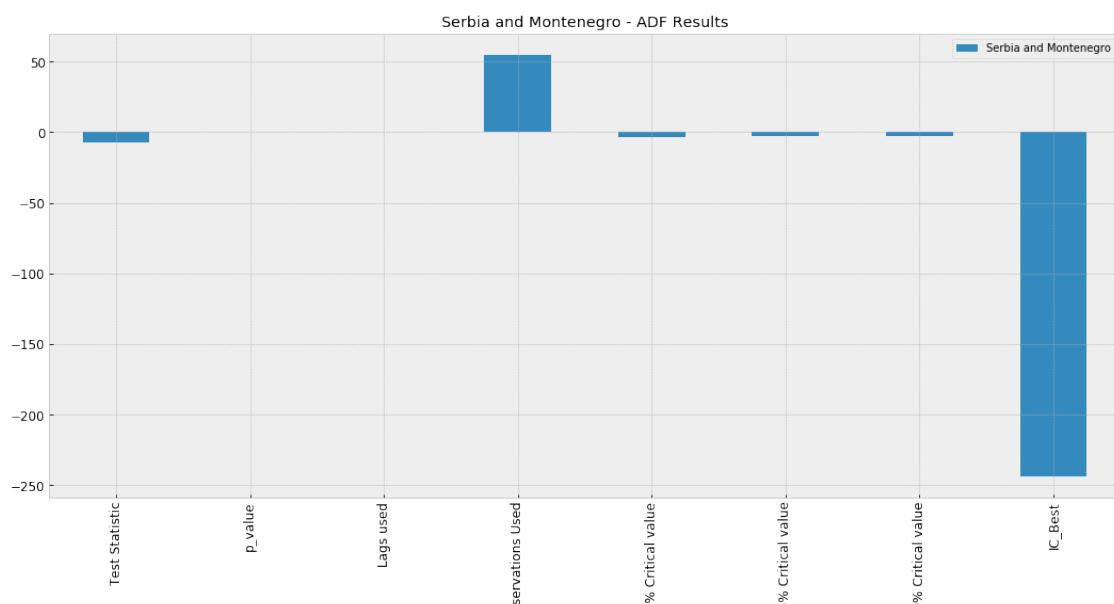
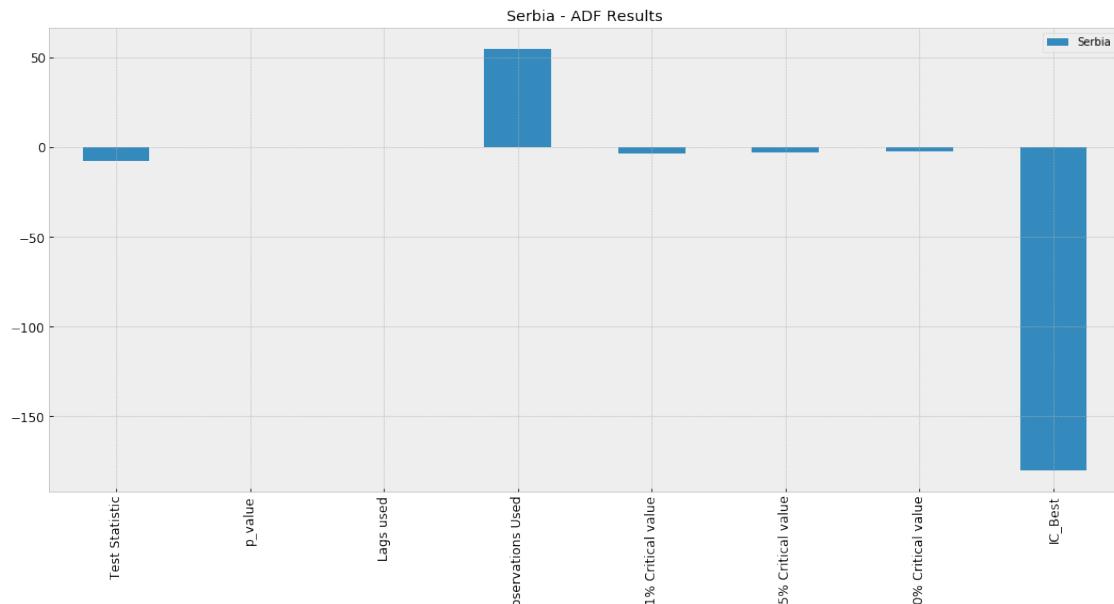
Saint Vincent and the Grenadines - ADF Results

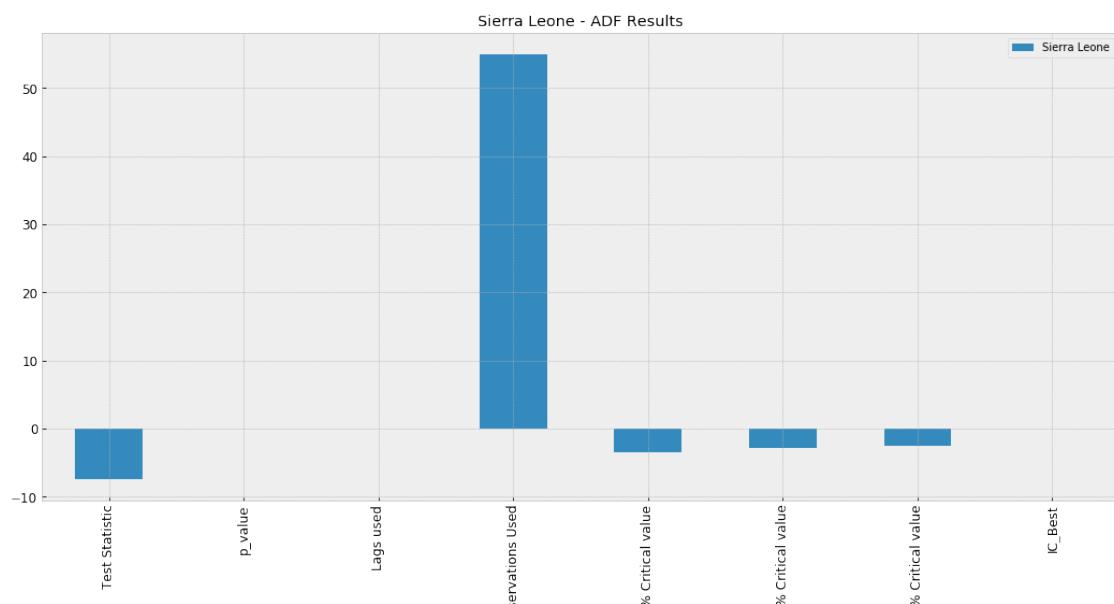
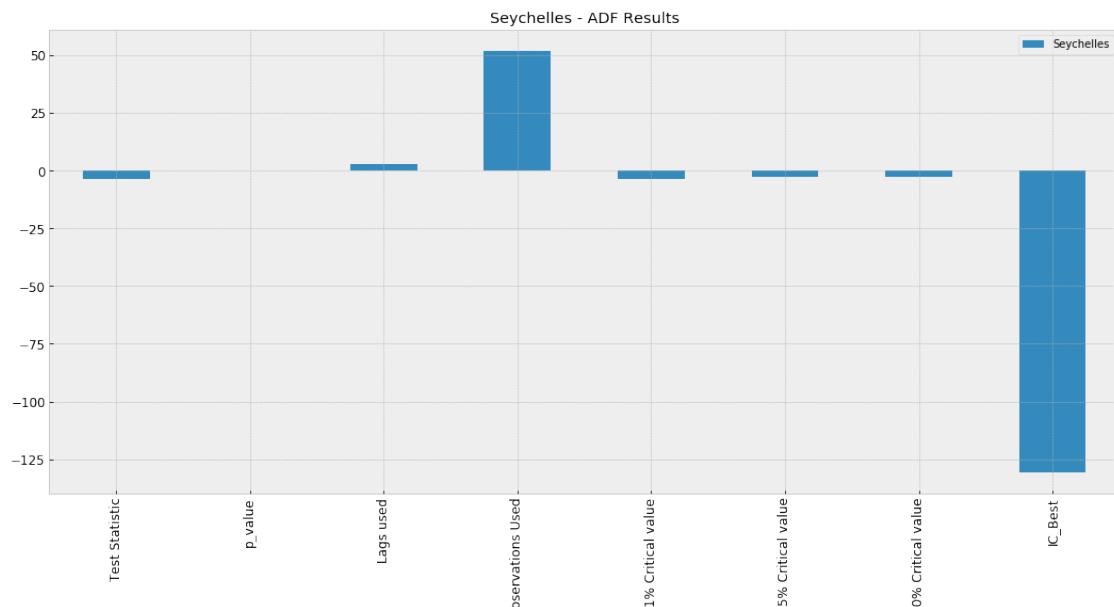


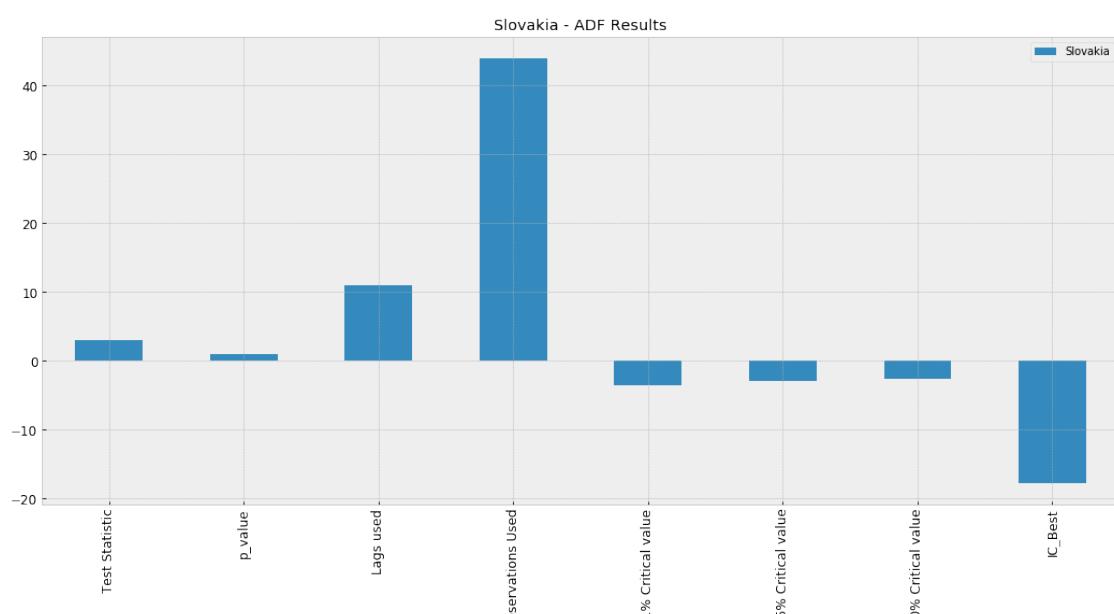
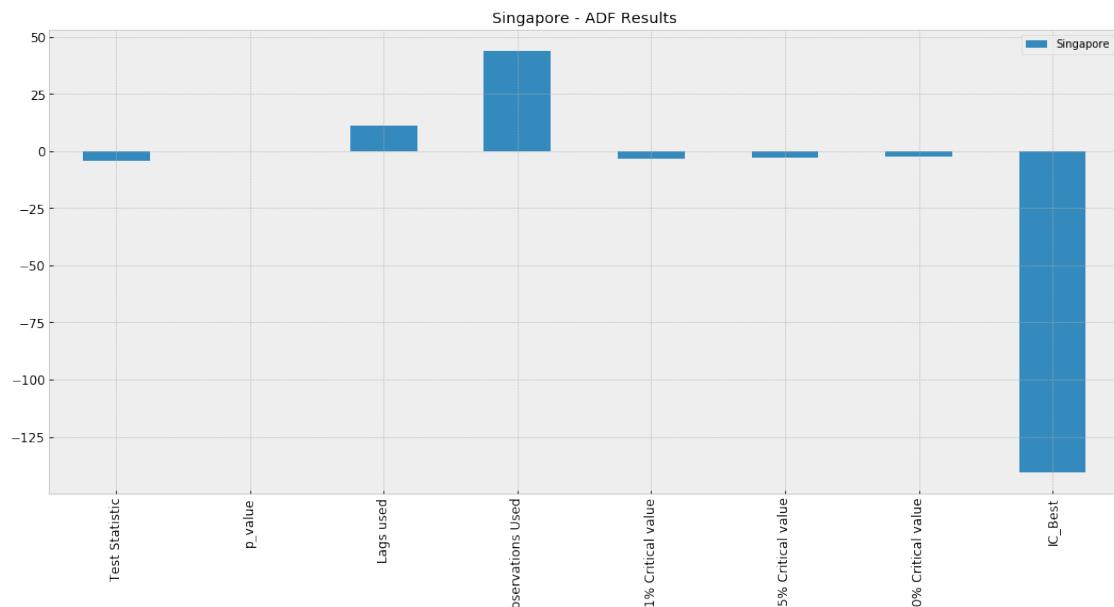
Samoa - ADF Results

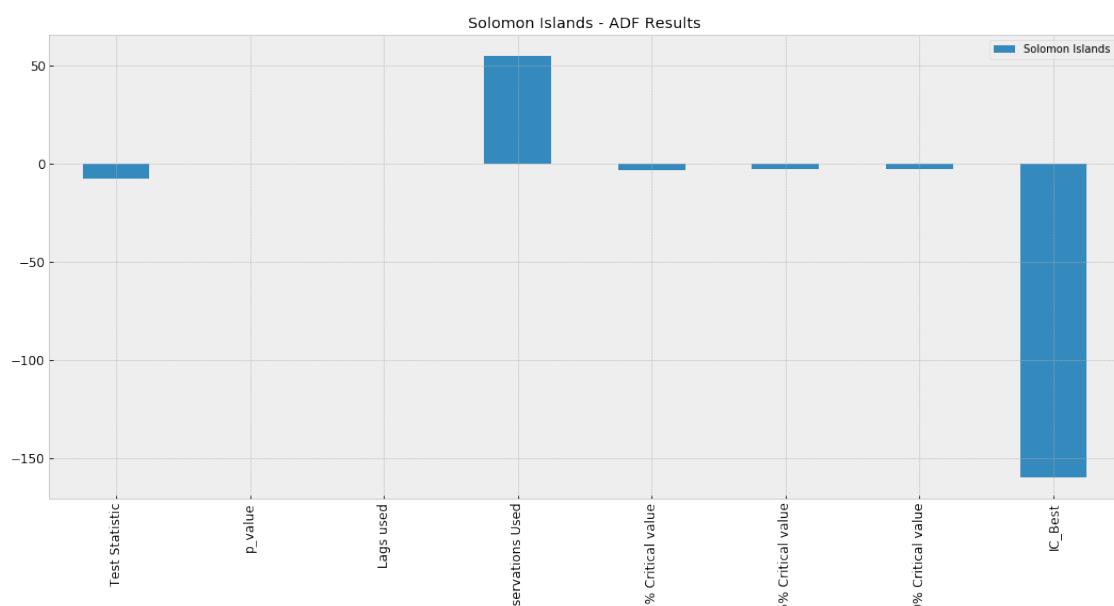
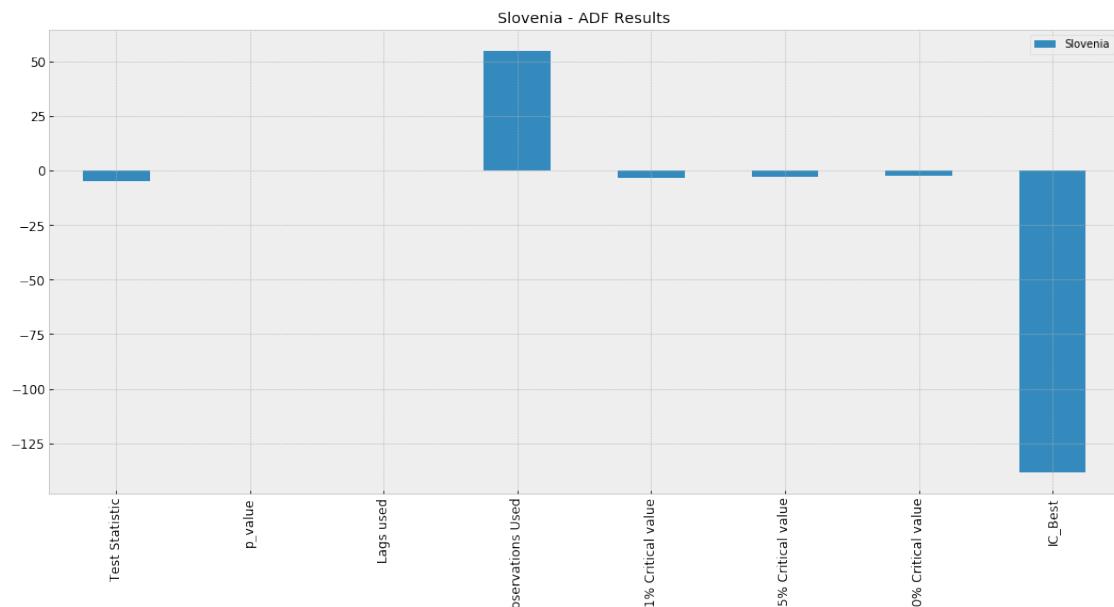


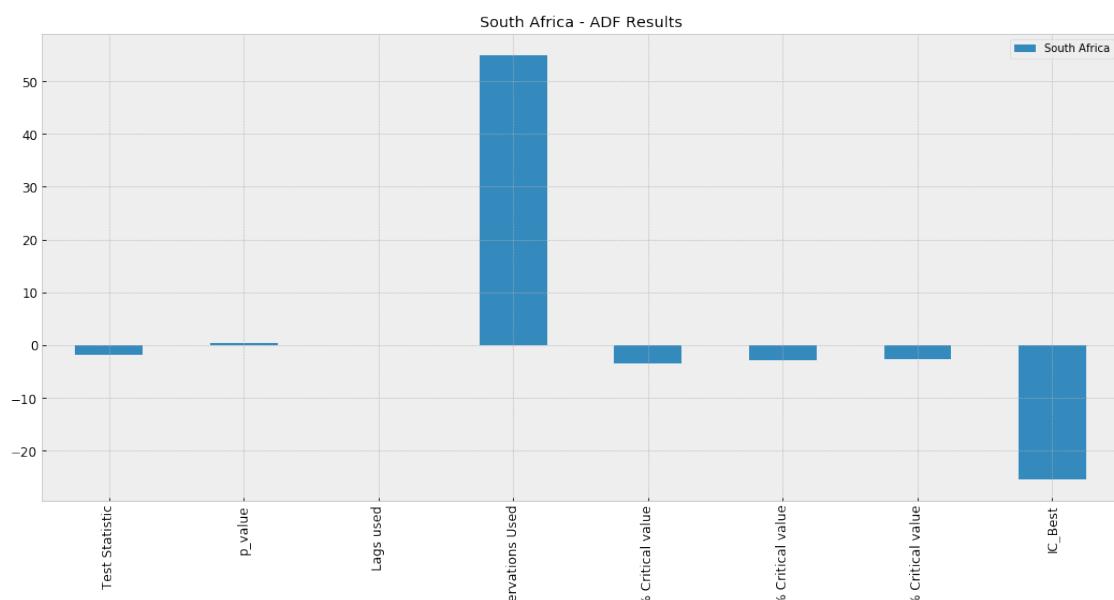
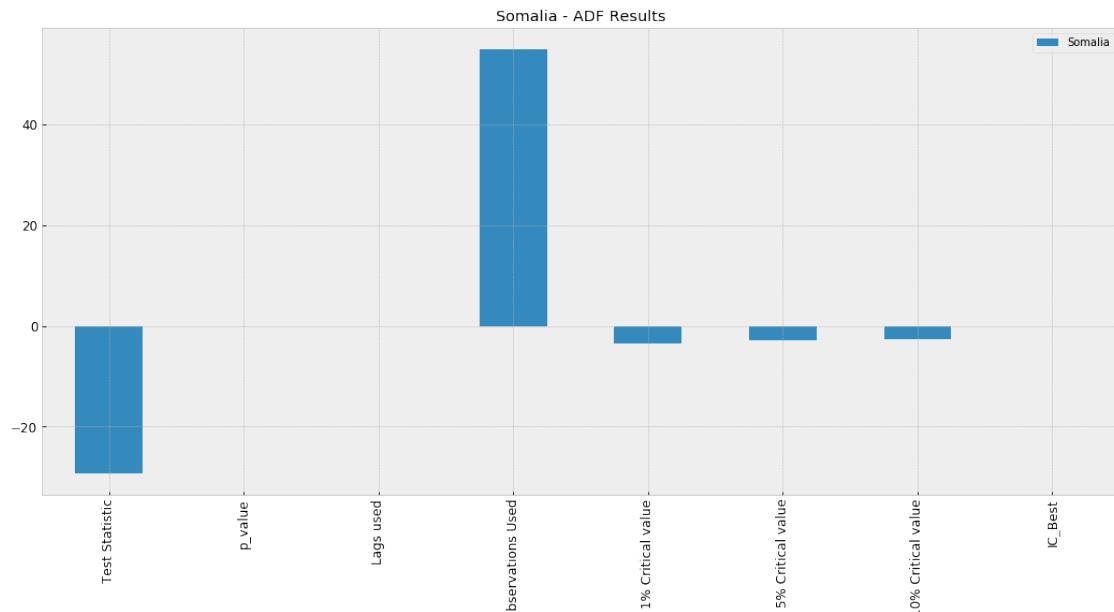


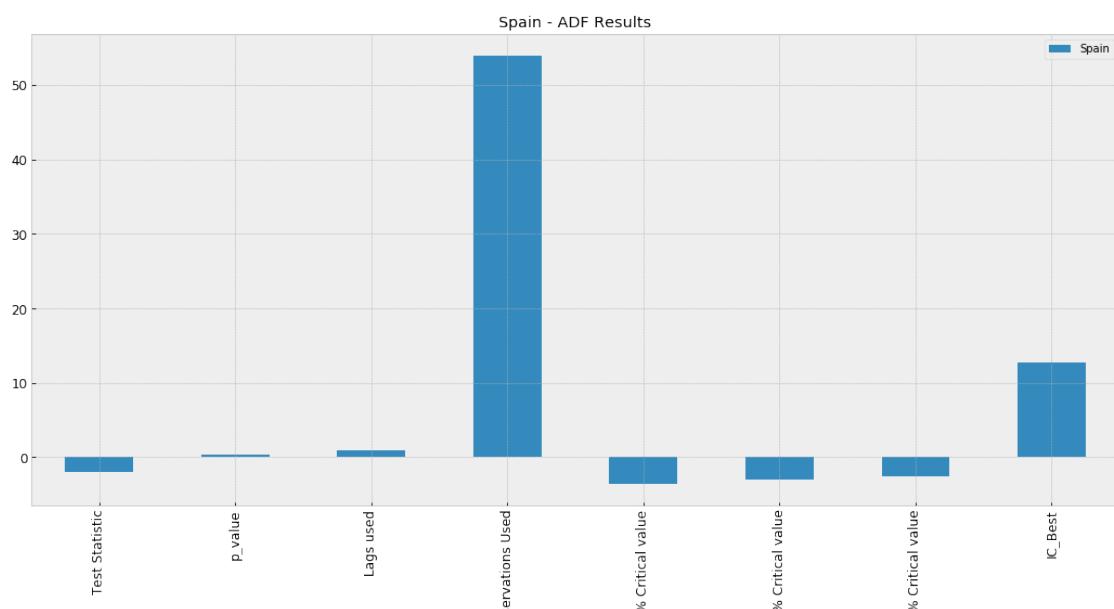
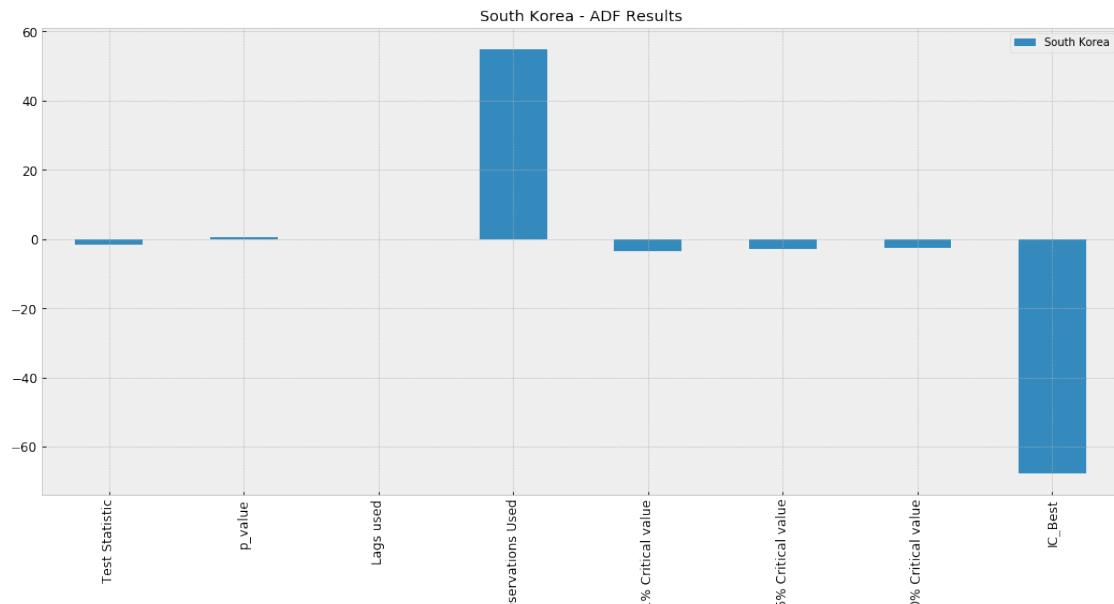


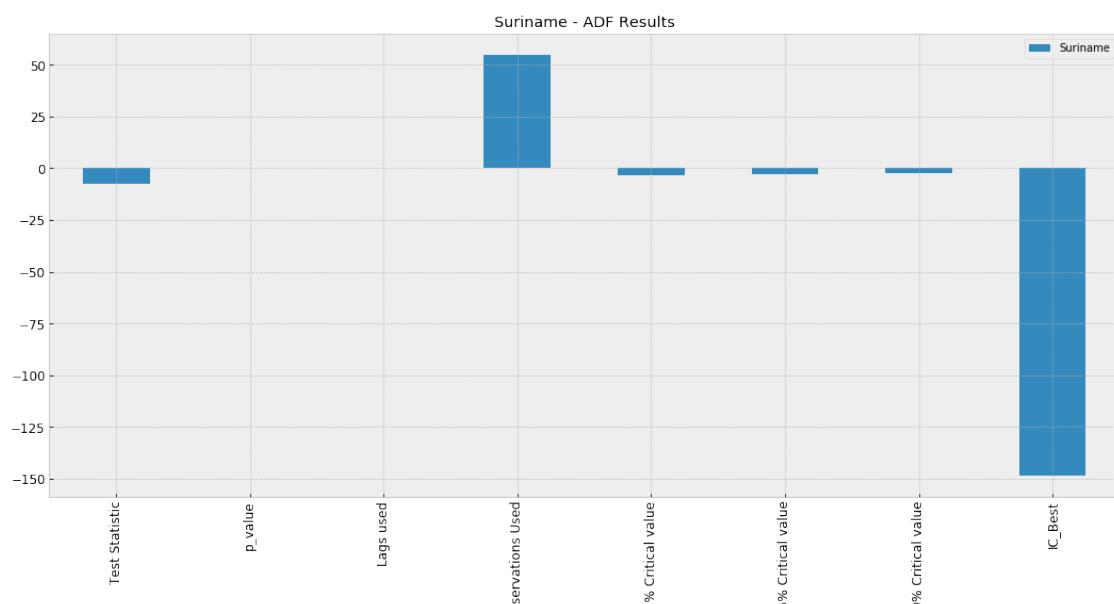
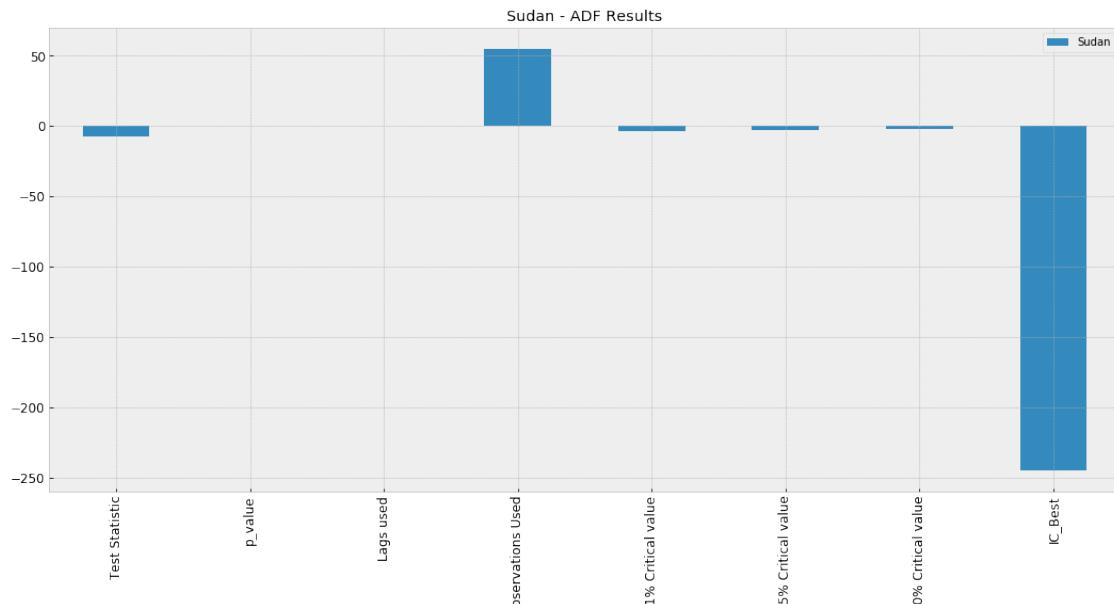


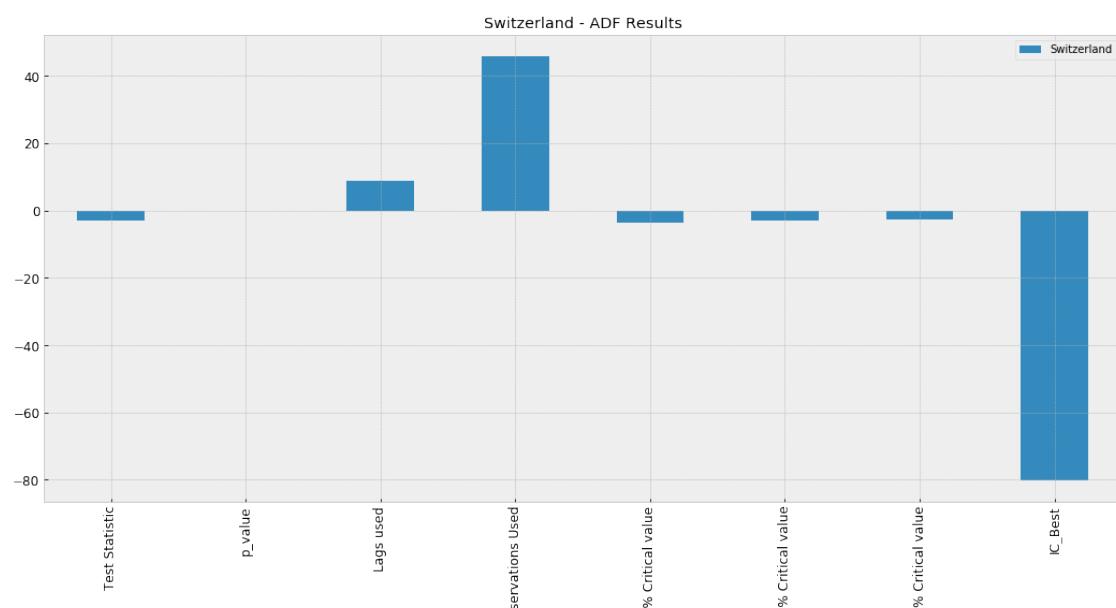
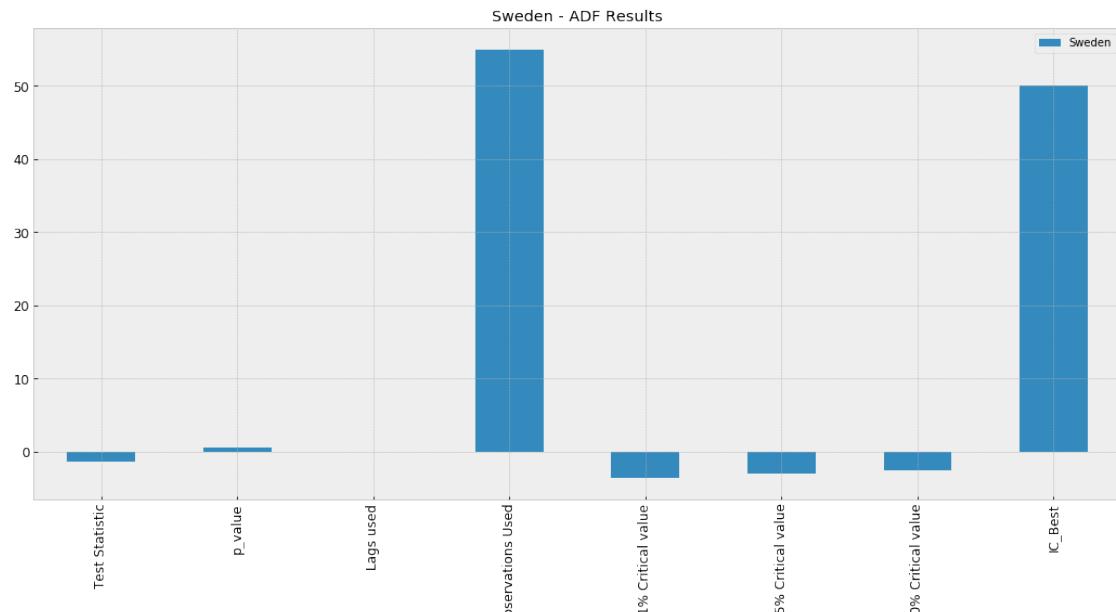


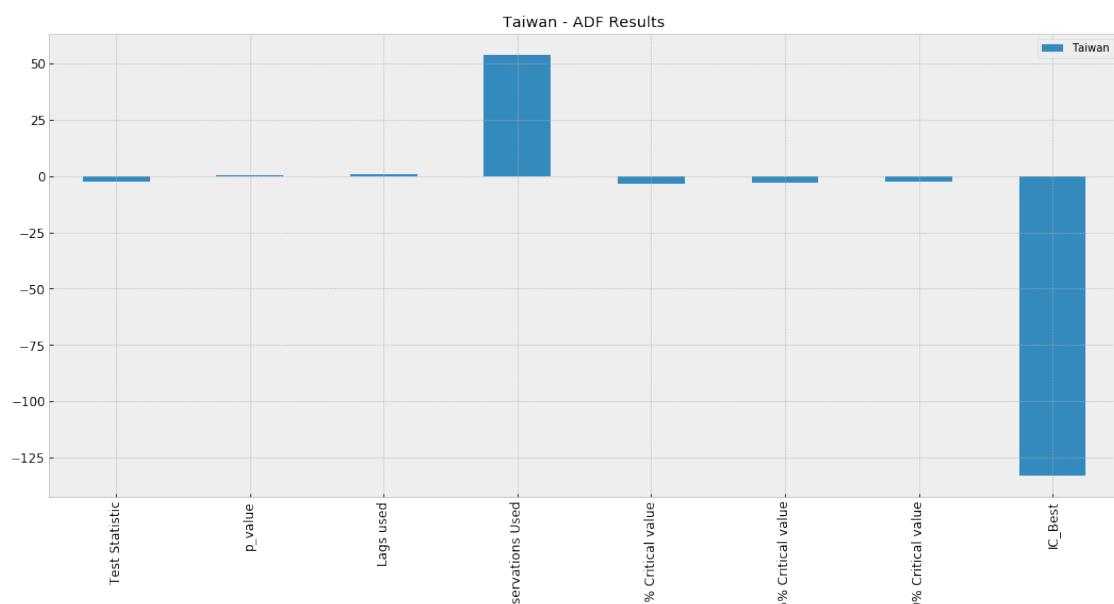
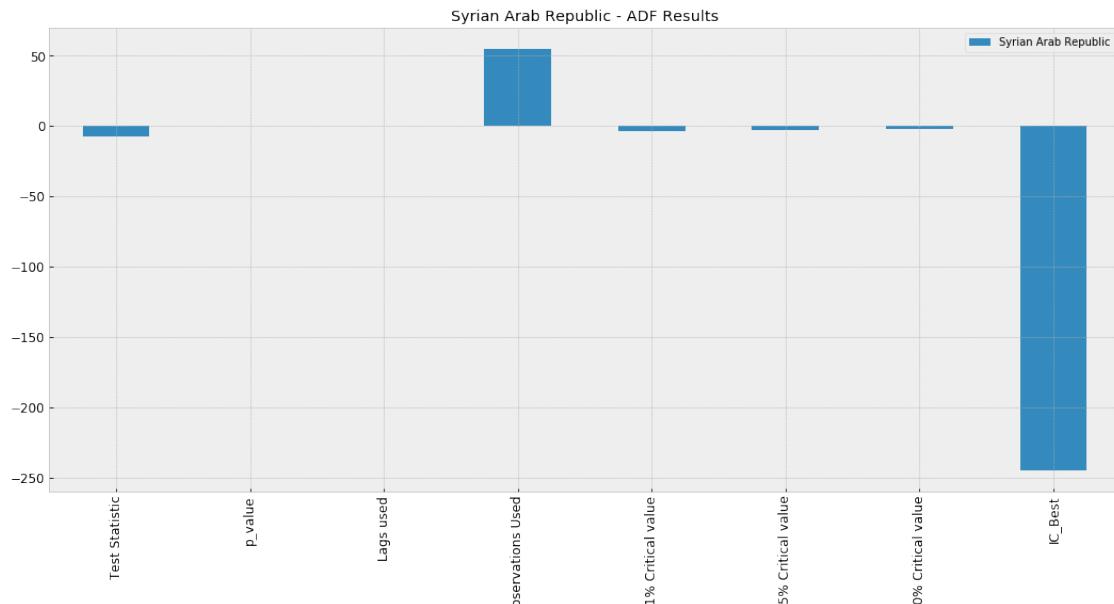


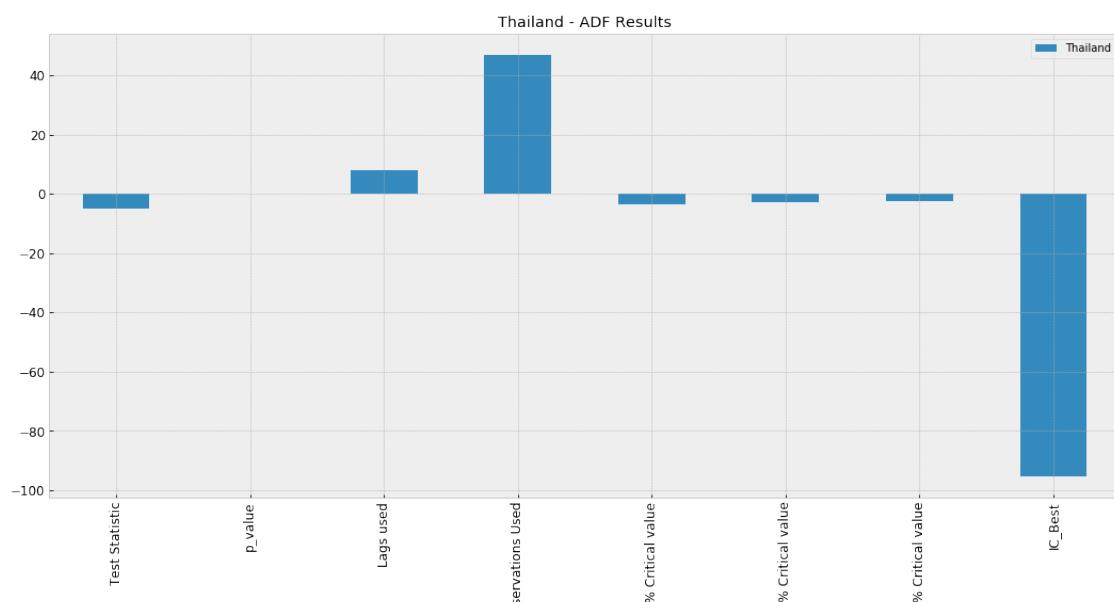
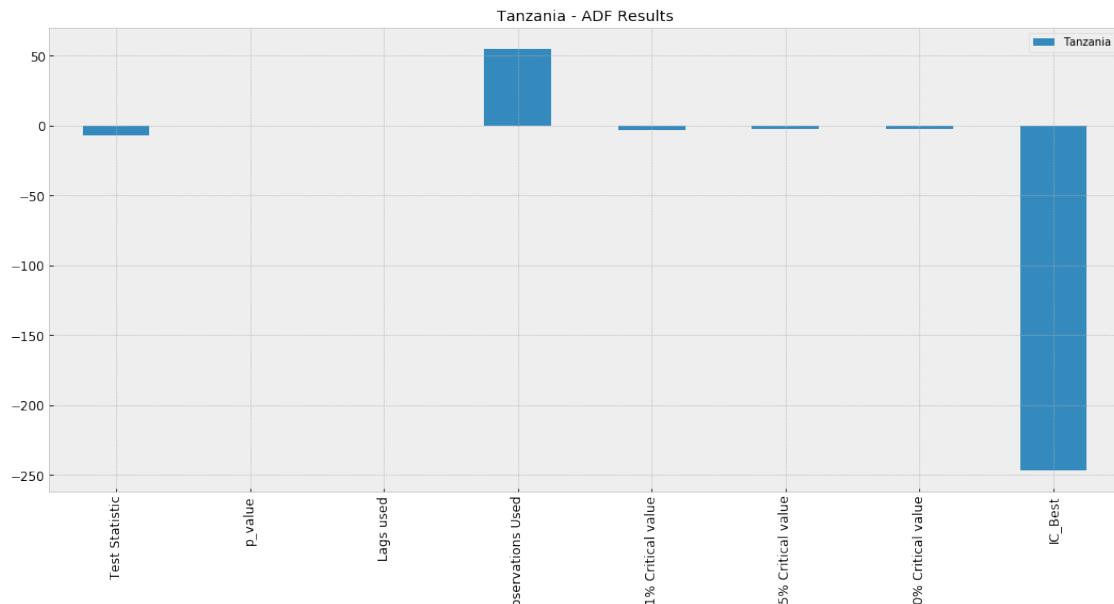


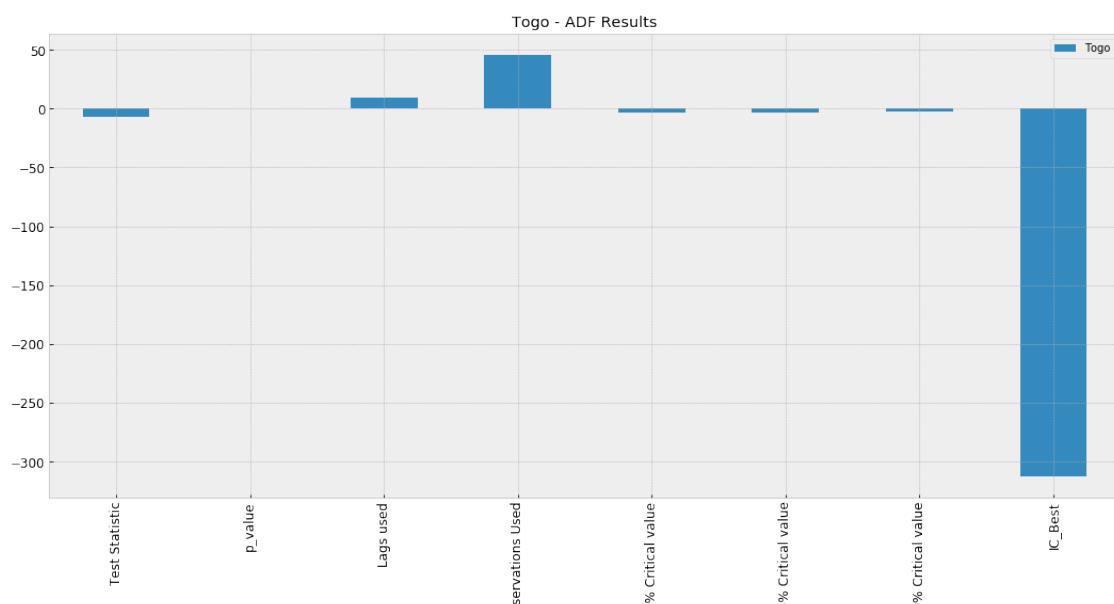
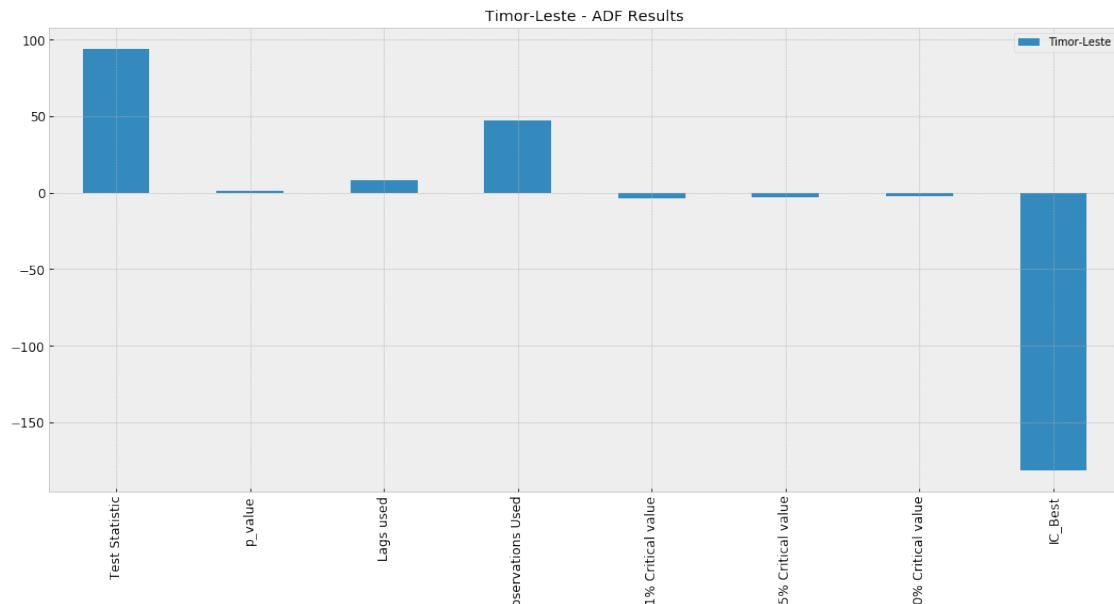


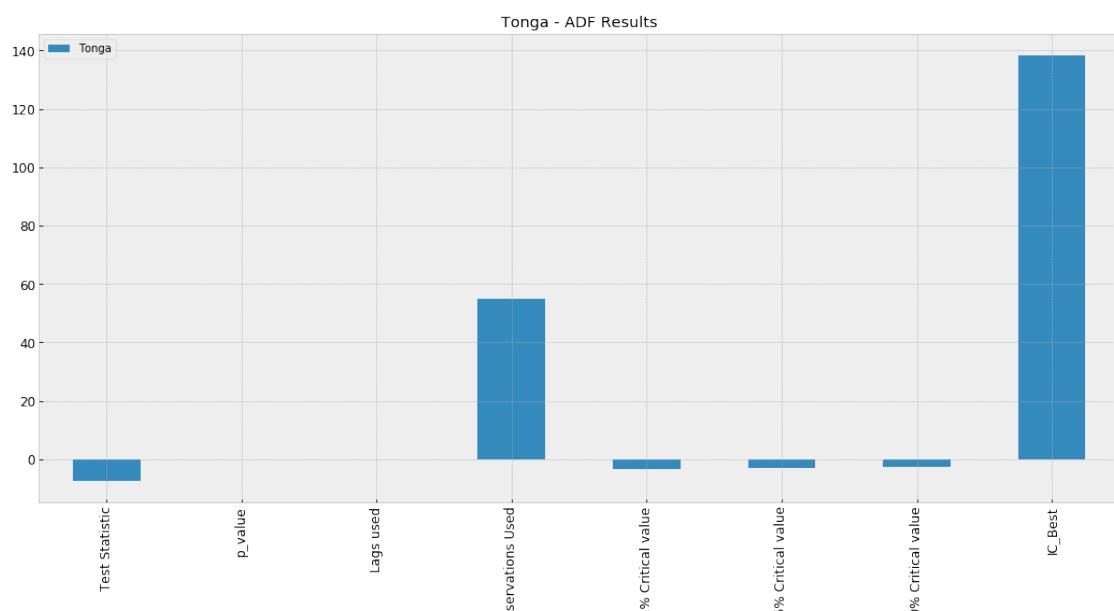
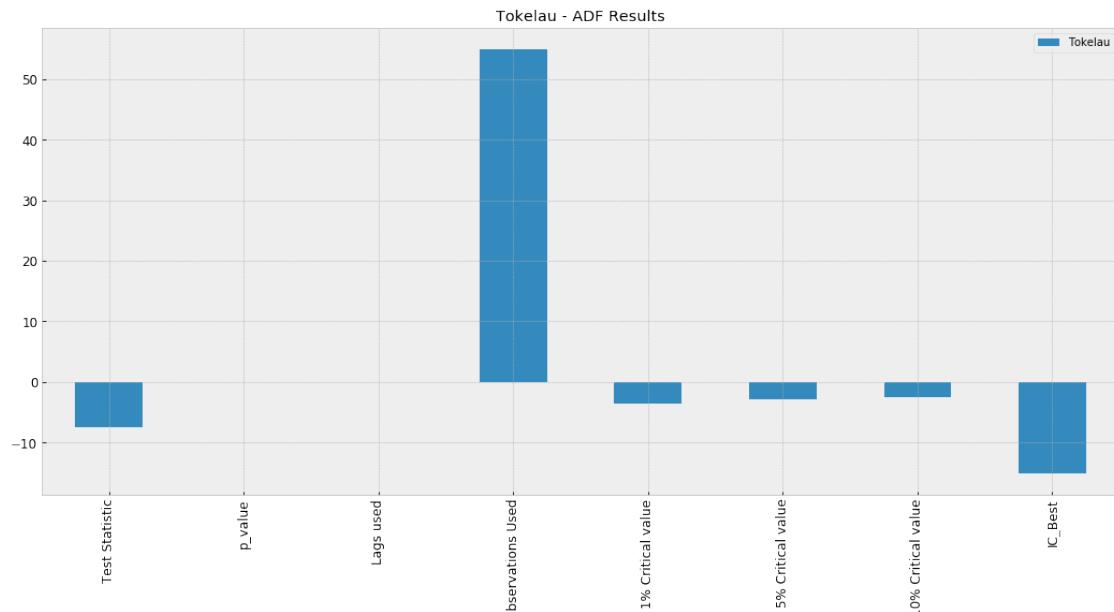


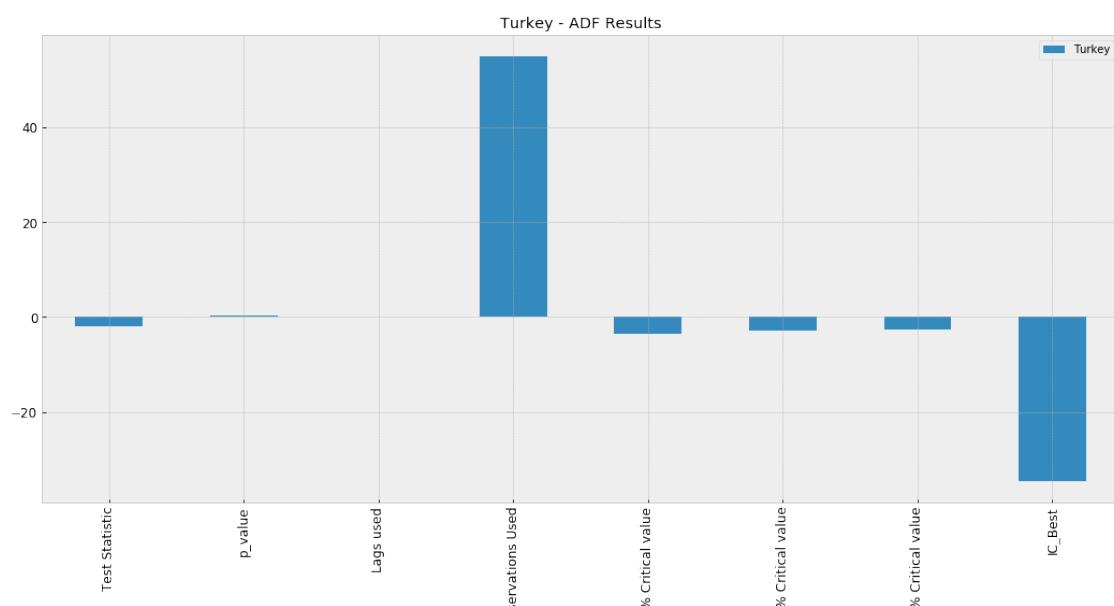
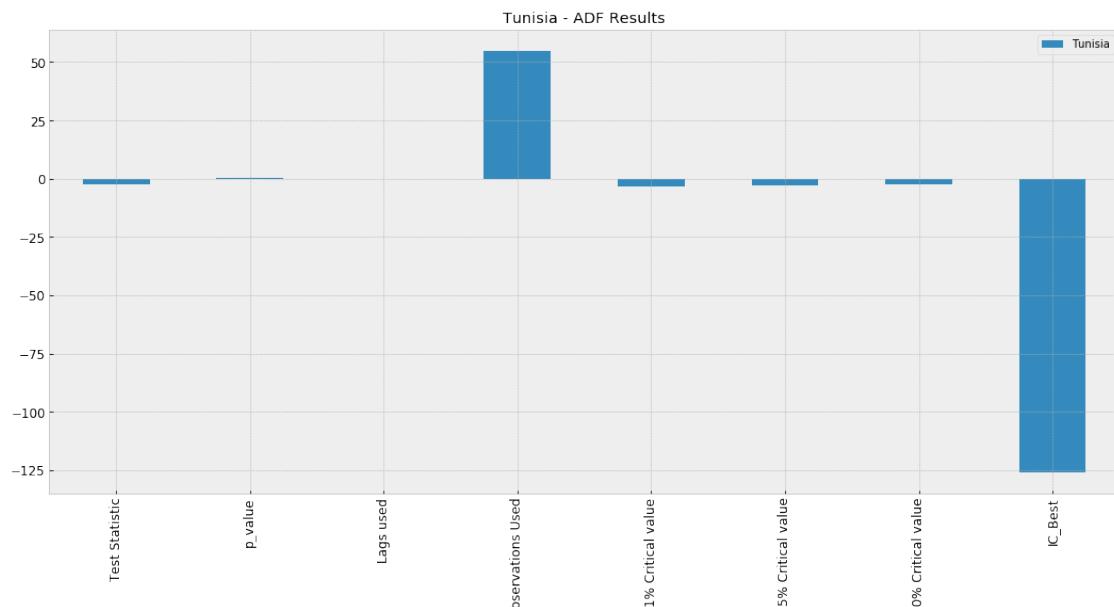


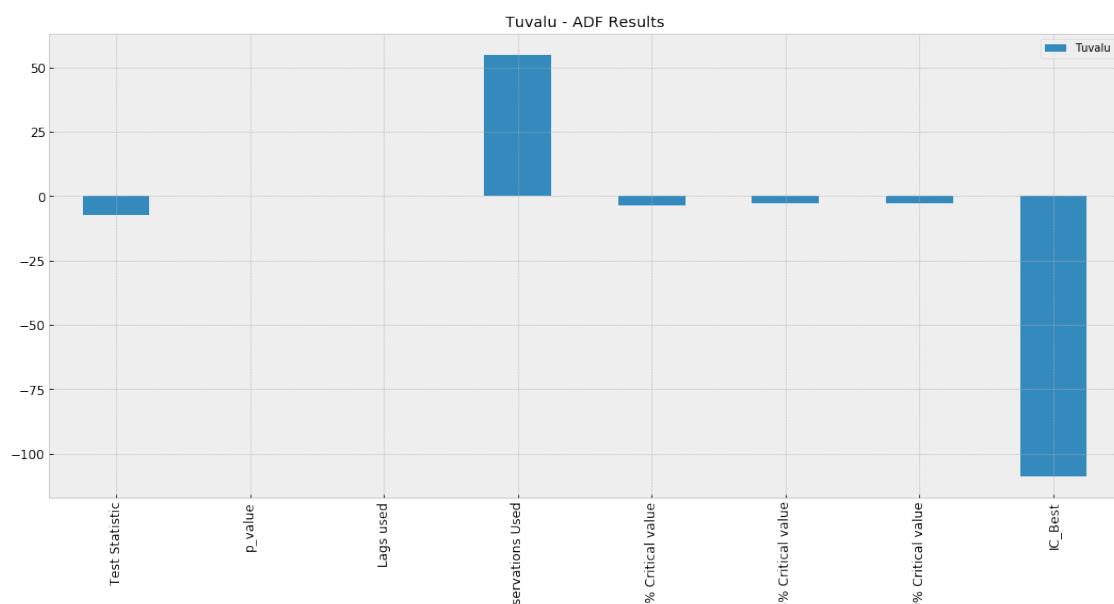
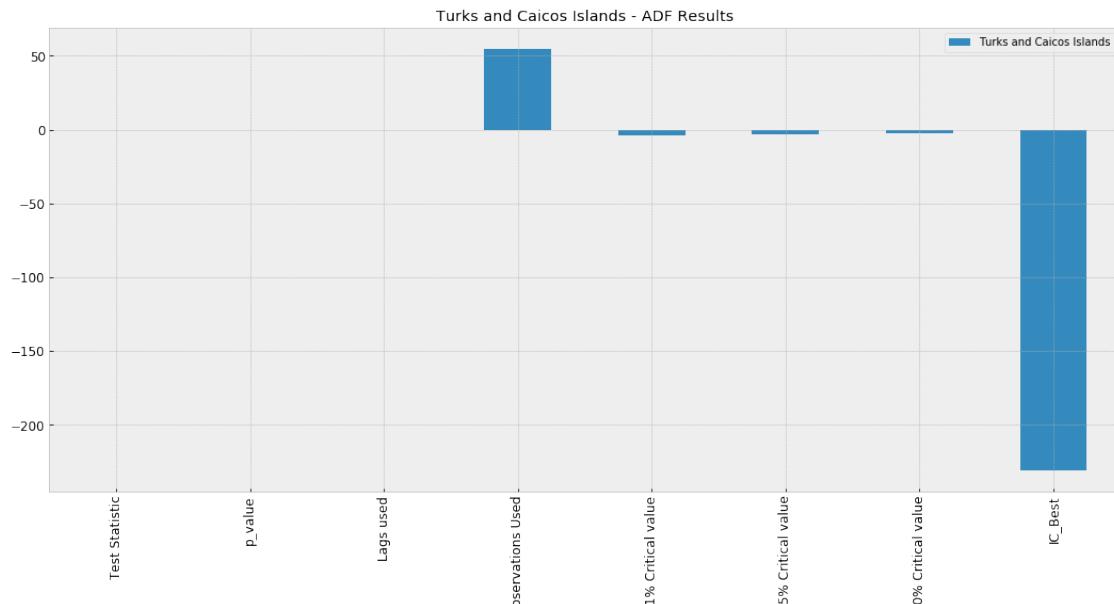


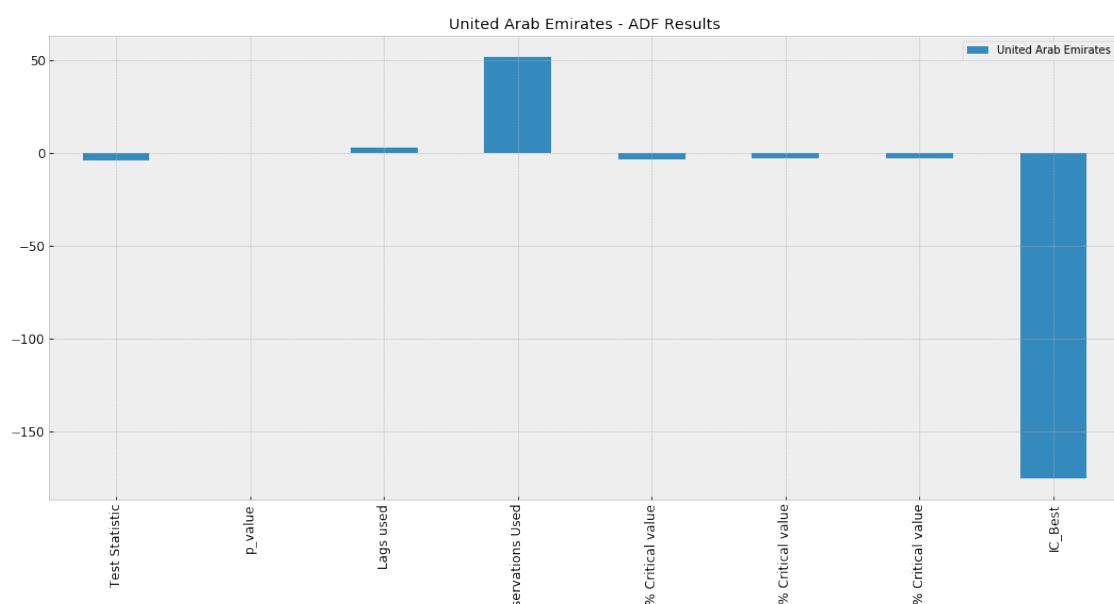
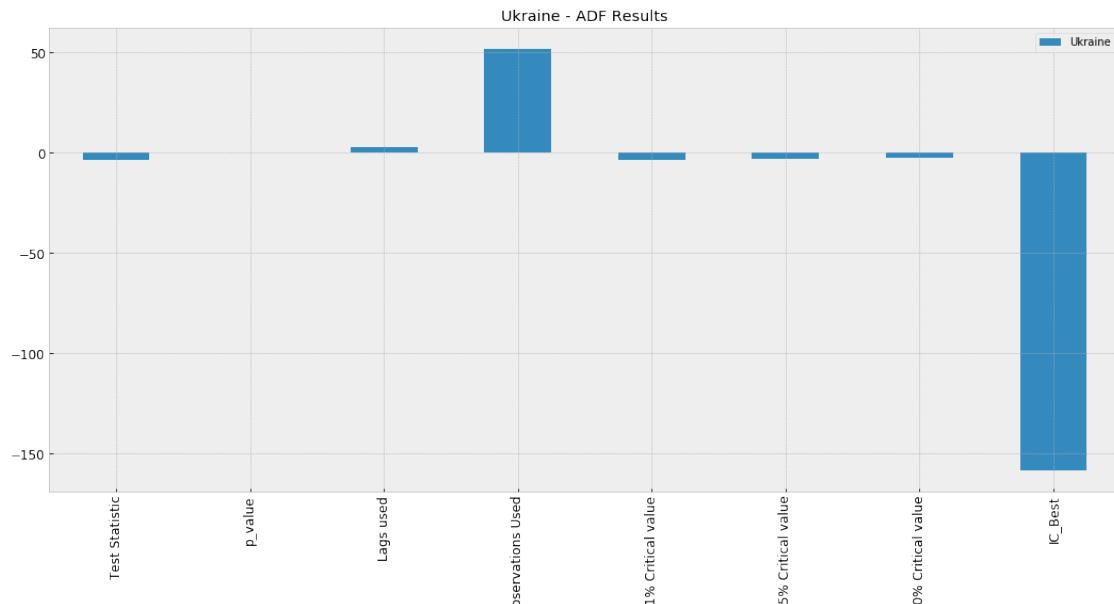


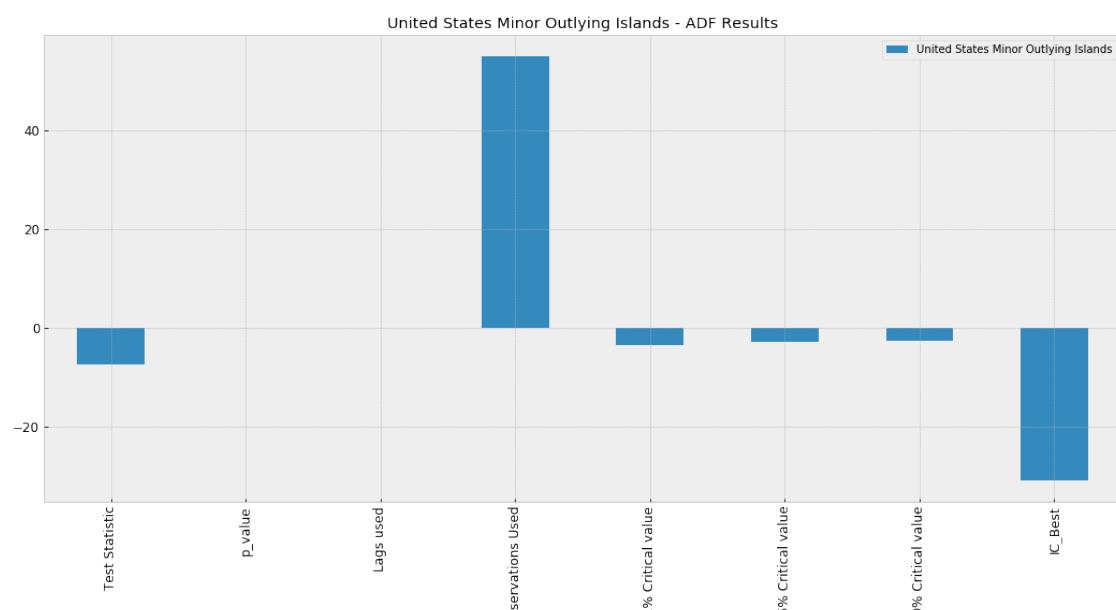
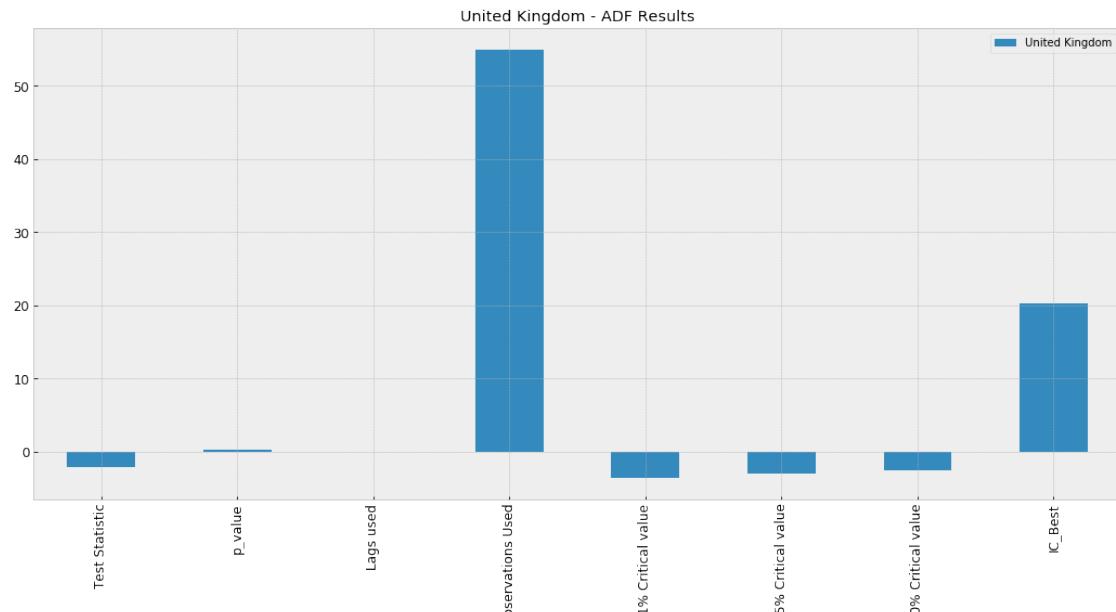


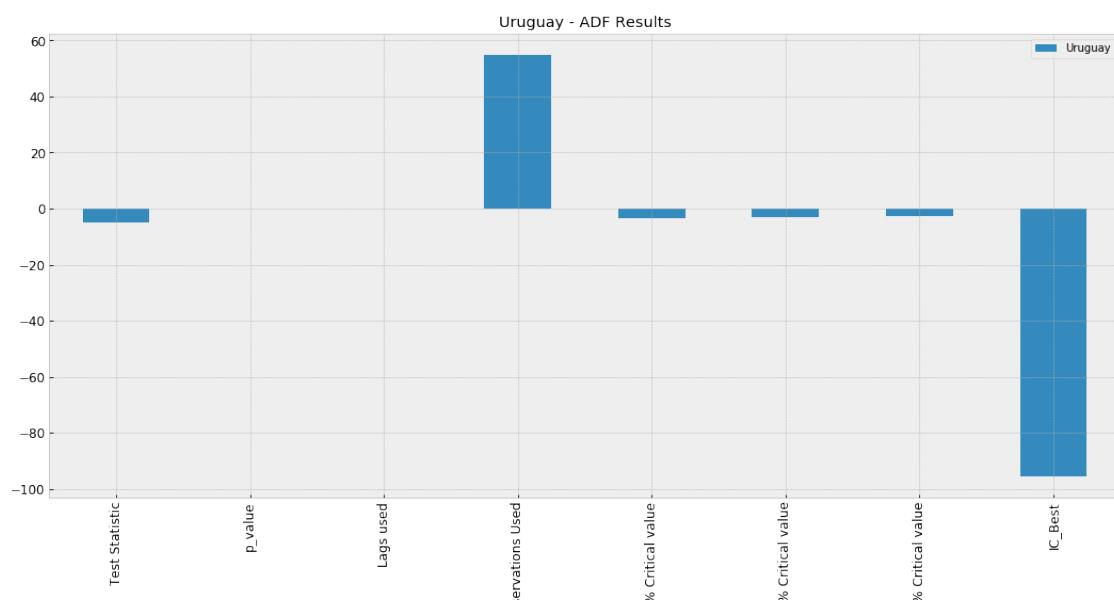
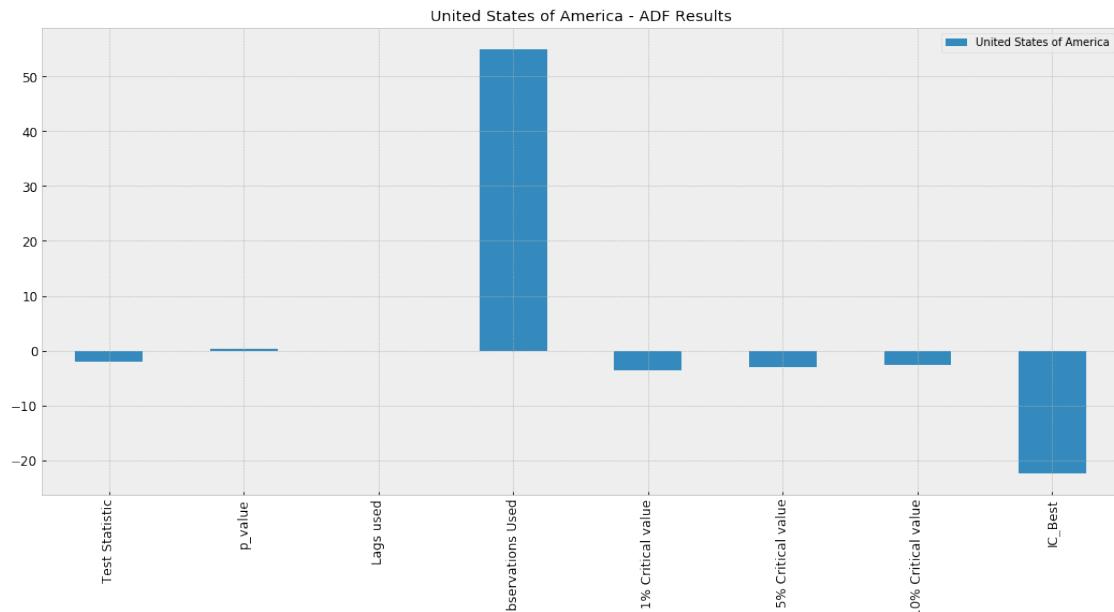


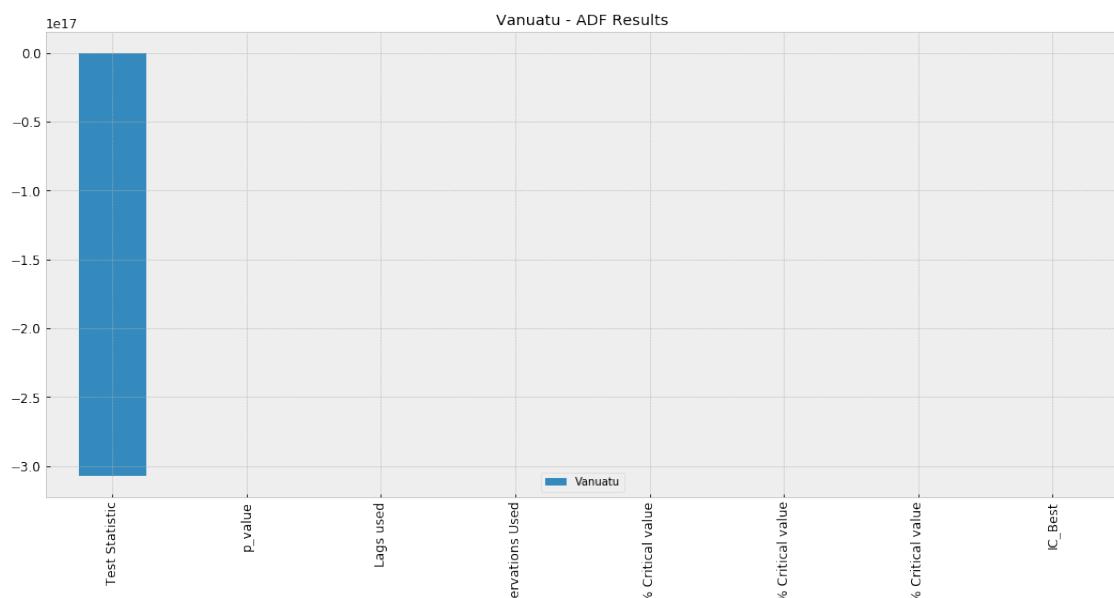
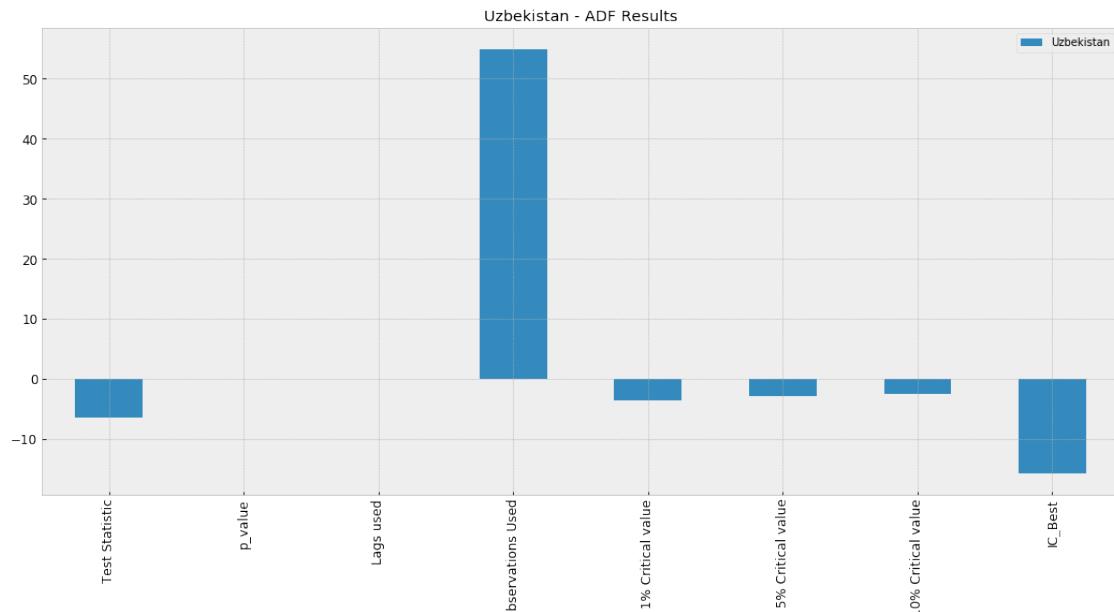


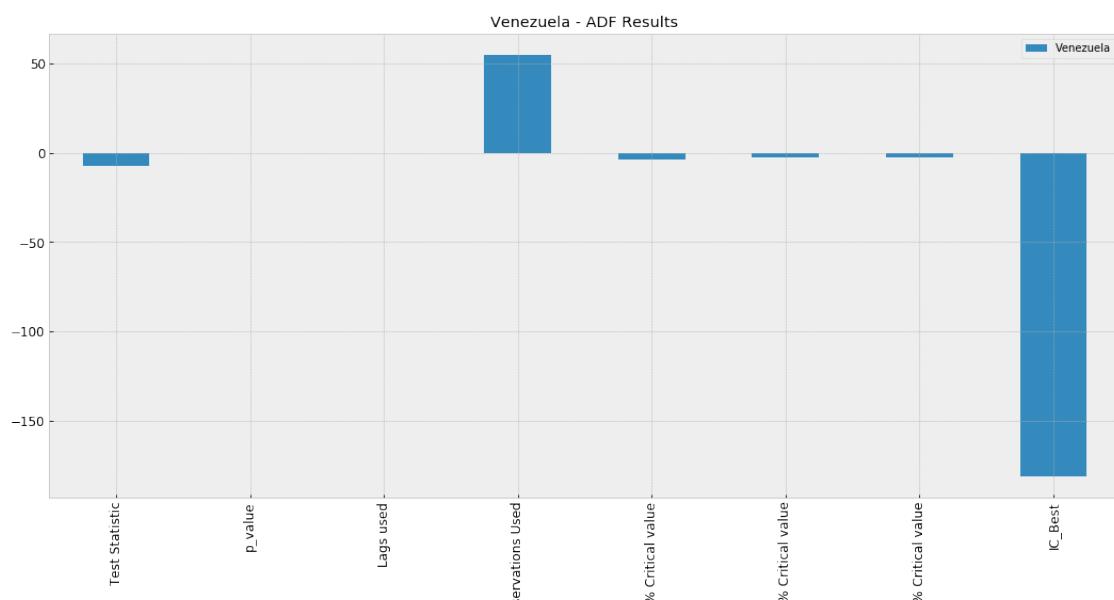
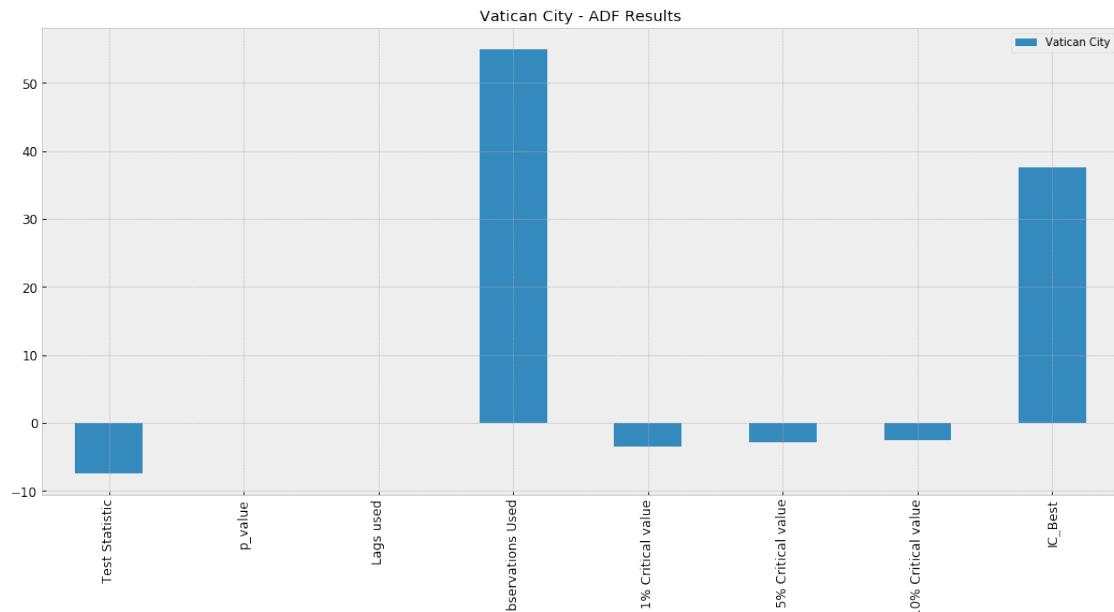


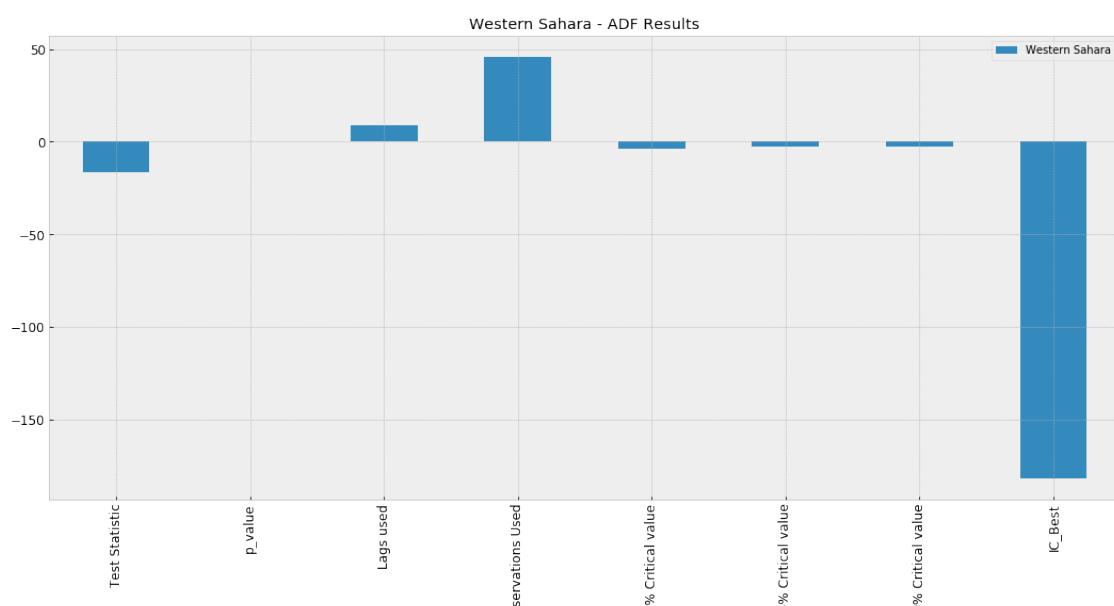
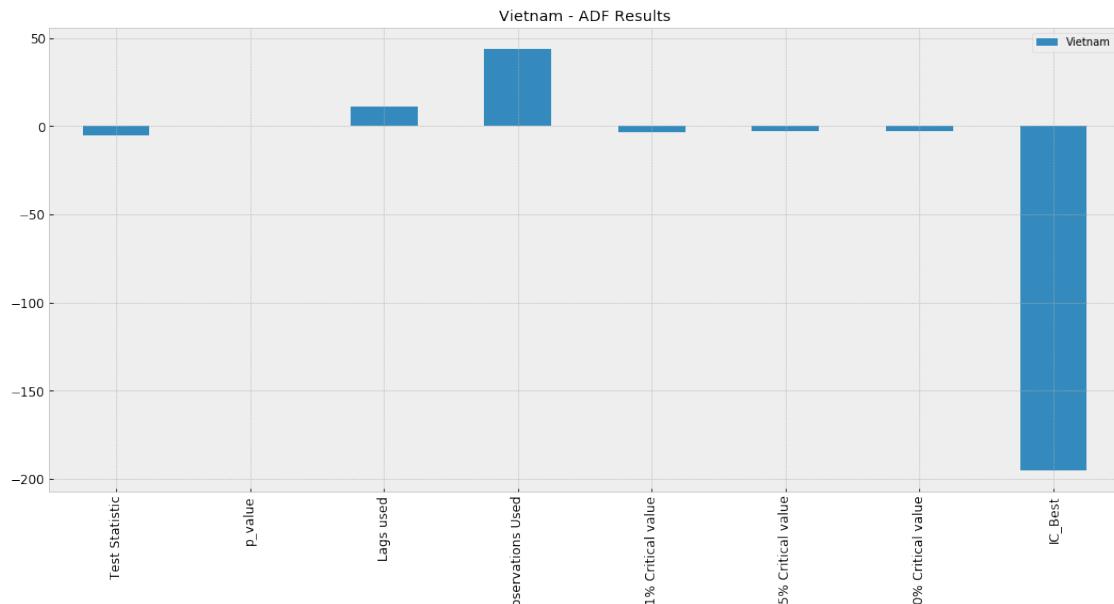


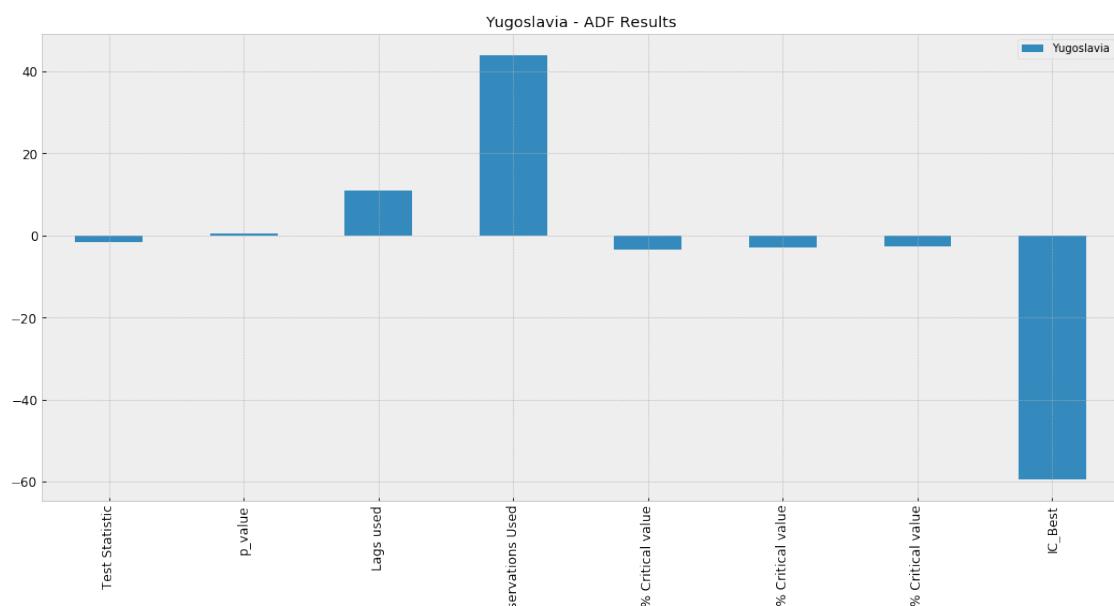
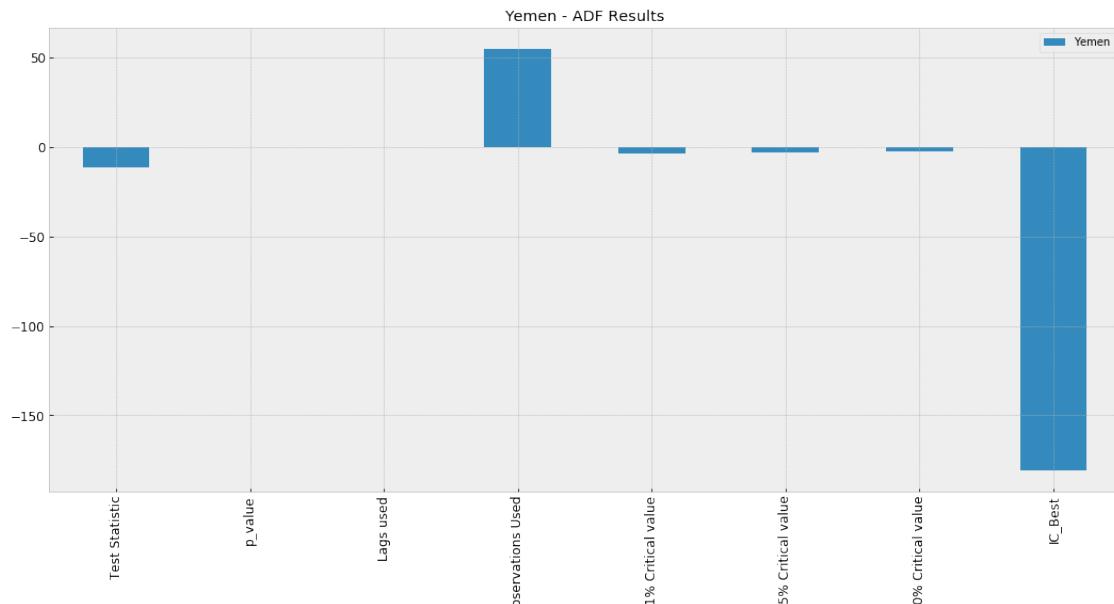


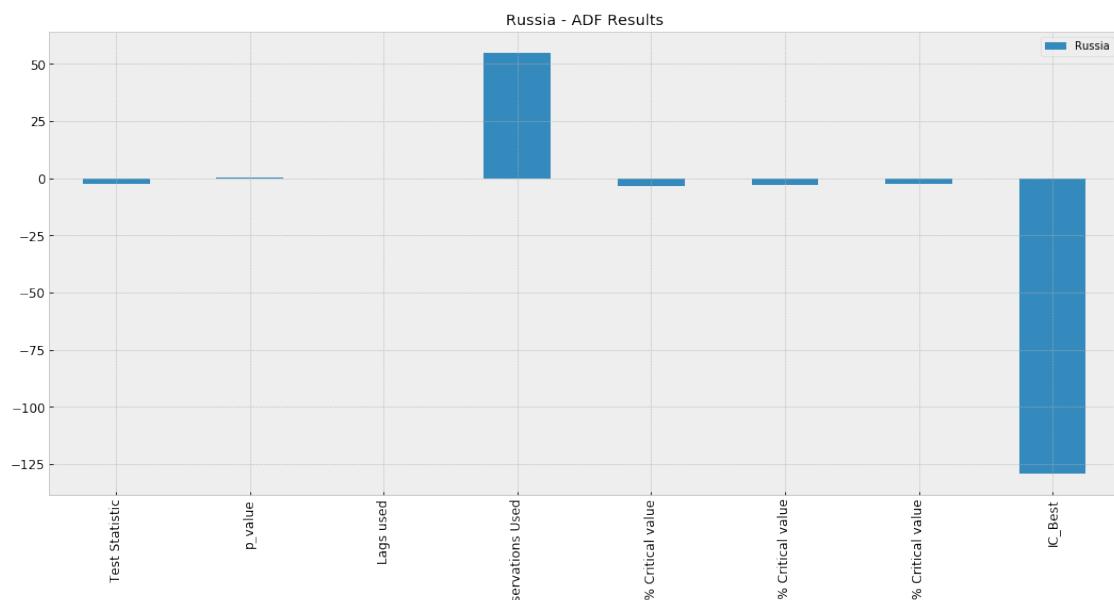
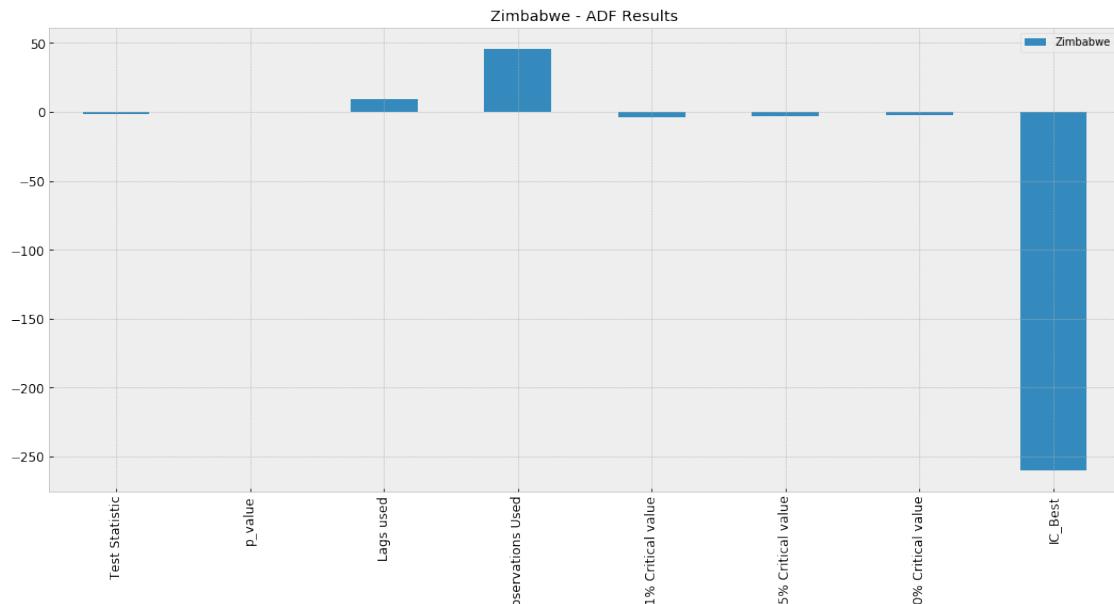












Automate p-value diagnosis: Try this to automate p-value rejection for non-stationary time series

4.2 Making Time Series Stationary

There are 2 major reasons behind non-stationaruty of a TS:

1. **Trend** – varying mean over time

2. **Seasonality** – variations at specific time-frames. eg people might have a tendency to buy cars in a particular month because of pay increment or festivals.

Our dataset had no seasonality, but suffers from a downward and an upward trend, which needs to be corrected in order for forecasting to be possible

[Other sources to check](#)

4.2.1 Transformations

- We can apply transformation which penalize higher values more than smaller values
- These can be taking a log, square root, cube root

Log Scale Transformation