Predicting Humanitarian Crises

Eleanor Alvarez

Onome Akpogheneta

Agenda

1

Introduction

Agenda

1. Introduction
2. Project 1
3. Project 3
4. Project 3

Agenda

1 Introduction

2 Project 1

3 Project 2

4 Project 3

1. Data cleaning & Exploratory Data Analysis
2. Machine Learning Methods
3. Machine Learning Results

Agenda

1	2	3	4	5
Introduction	Project 1	Project 2	Project 3	Conclusions

1. Data cleaning & Exploratory Data Analysis
2. Machine Learning Methods
3. Machine Learning Results

Back to Agenda

Introduction

What is a humanitarian crisis?

Back to Agenda

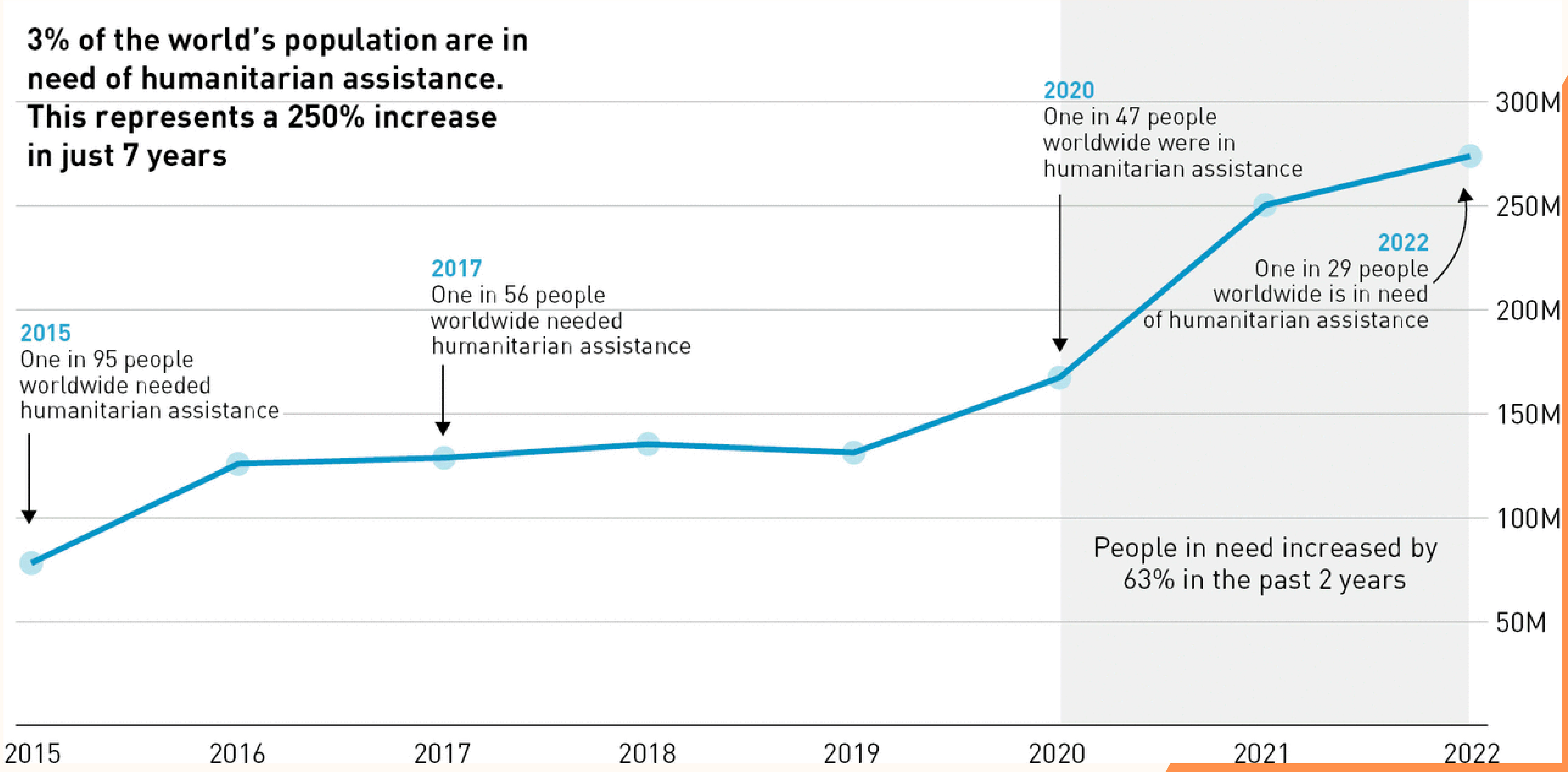
A situation or event where there is a ***significant threat to the health, safety, and well-being of a large number of people****, often resulting from factors such as natural disasters, armed conflict, or socio-political instability. This can lead to widespread* ***displacement****,* ***loss of life****, and* ***critical shortages*** *of basic necessities such as food, water, shelter, and medical care*

Source: Global Humanitarian Overview 2022. United Nations Office for the Coordination of Humanitarian Affairs

What is a humanitarian crisis?

Back to Agenda

A situation or event where there is a ***significant threat to the health, safety, and well-being of a large number of people****, often resulting from factors such as natural disasters, armed conflict, or socio-political instability. This can lead to widespread* ***displacement****,* ***loss of life****, and* ***critical shortages*** *of basic necessities such as food, water, shelter, and medical care*



Source: Global Humanitarian Overview 2022. United Nations Office for the Coordination of Humanitarian Affairs

Humanitarian crises can be prevented

Back to Agenda

The humanitarian sector is primarily focused on **reponse**

If our approach was more **proactive** and **anticipatory** than countless lives could be saved

EXAMPLES

Early Action

African Union Insurance

Intro to the Dataset

Back to Agenda

INFORM

SHARING CRISIS ANALYSIS

European Commission

INFORM is a multi-stakeholder forum for developing **shared, quantitative analysis relevant to humanitarian crises and disasters**

INFORM is developing a suite of products to **support decision-making** on humanitarian crises and disasters

These products help make decisions at different stages of the disaster management cycle, specifically **climate adaptation** and **disaster prevention**, **preparedness and response**

1. **INFORM RISK**
2. **INFORM CLIMATE CHANGE**
3. **INFORM SEVERITY**

Can humanitarian crises be predicted?

Back to Agenda

Project 1

Explore whether the INFORM data products can be used to predict whether or not there is an ongoing humanitarian crisis

Can humanitarian crises be predicted?

Back to Agenda

Project 1

Explore whether the INFORM data products can be used to predict whether or not there is an ongoing humanitarian crisis

Project 2

Explore whether the INFORM data products can be used to predict the severity of humanitarian crises

Can humanitarian crises be predicted?

Back to Agenda

Project 1

Explore whether the INFORM data products can be used to predict whether or not there is an ongoing humanitarian crisis

Project 2

Explore whether the INFORM data products can be used to predict the severity of humanitarian crises

Project 3

Predict & forecast how INFORM risk data changes over time

Back to Agenda

Project 1

Explore whether the INFORM data products can be used to predict whether or not there is an ongoing humanitarian crisis

Data Cleaning & EDA

INFORM Risk Dataset

Back to Agenda

COUNTRY (a-z)	ISO3 (a-z)	INFORM RISK (0-10)	RISK CLASS (Very Low-Very High)
Afghanistan	AFG	8.1	Very High
Albania	ALB	3.1	Low
Algeria	DZA	3.6	Medium
Angola	AGO	5.2	High
Antigua and Barbuda	ATG	2.3	Low
Argentina	ARG	2.9	Low
Armenia	ARM	2.9	Low
Australia	AUS	2.4	Low
Austria	AUT	1.9	Very Low
Azerbaijan	AZE	4.8	Medium
Bahamas	BHS	2.1	Very Low
Bahrain	BHR	1.4	Very Low
Bangladesh	BGD	5.7	High

INFORM Risk Dataset

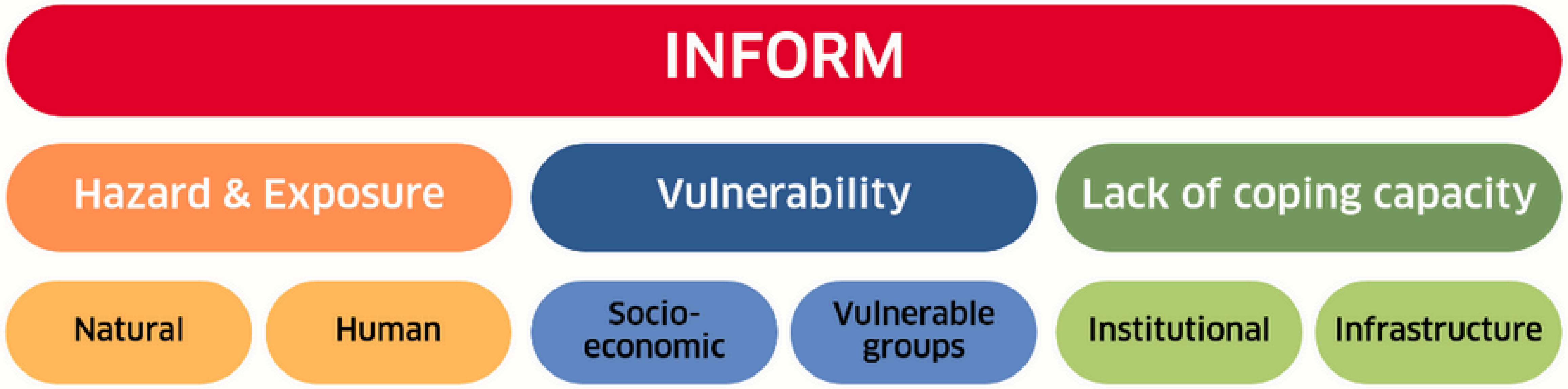
Back to Agenda

COUNTRY (a-z)	ISO3 (a-z)	INFORM RISK (0-10)	RISK CLASS (Very Low-Very High)
Afghanistan	AFG	8.1	Very High
Albania	ALB	3.1	Low
Algeria	DZA	3.6	Medium
Angola	AGO	5.2	High
Antigua and Barbuda	ATG	2.3	Low
Argentina	ARG	2.9	Low
Armenia	ARM	2.9	Low
Australia	AUS	2.4	Low
Austria	AUT	1.9	Very Low
Azerbaijan	AZE	4.8	Medium
Bahamas	BHS	2.1	Very Low
Bahrain	BHR	1.4	Very Low
Bangladesh	BGD	5.7	High

INFORM Risk Dataset

Back to Agenda

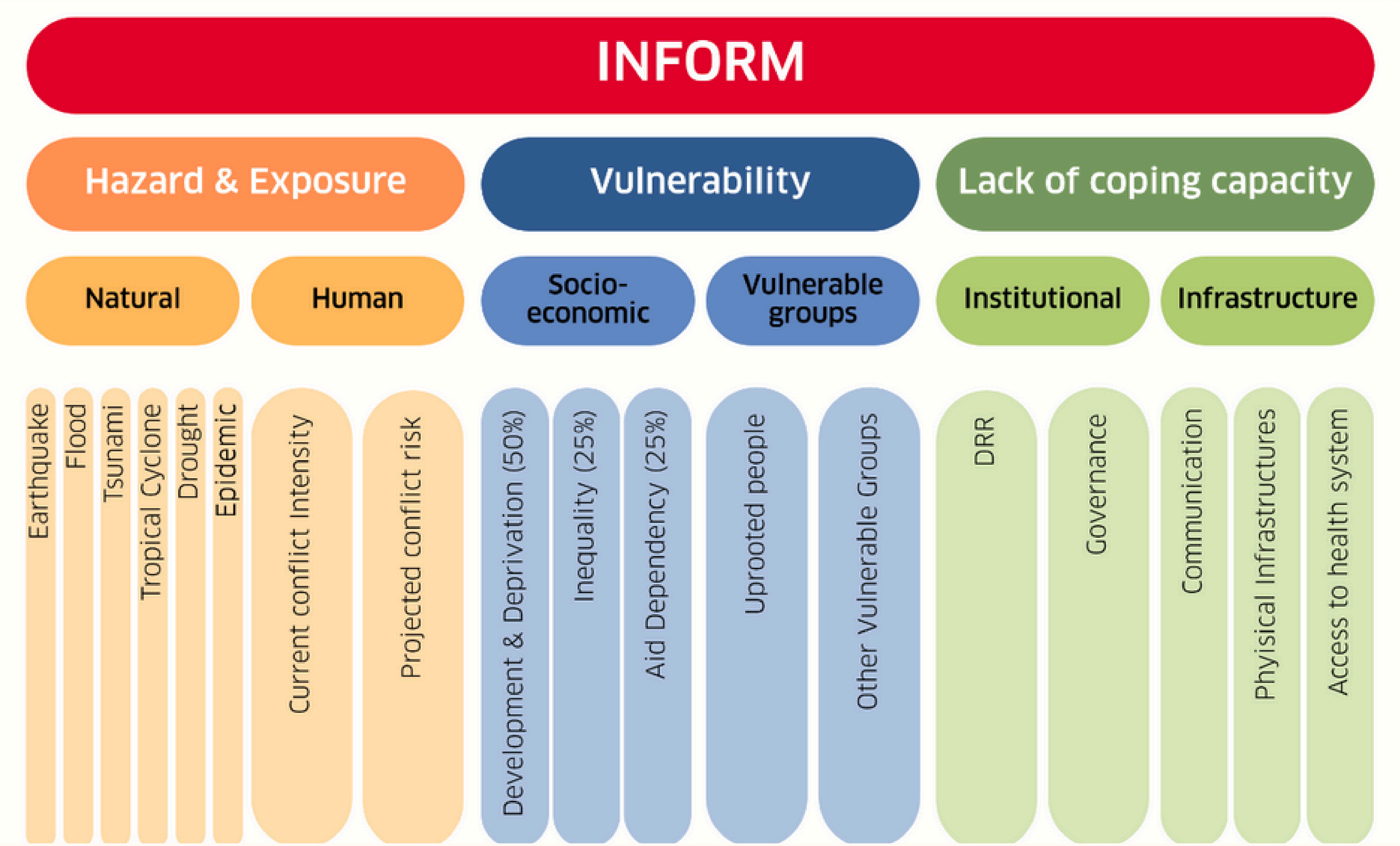
COUNTRY (a-z)	ISO3 (a-z)	INFORM RISK (0-10)	RISK CLASS (Very Low-Very High)
Afghanistan	AFG	8.1	Very High
Albania	ALB	3.1	Low
Algeria	DZA	3.6	Medium
Angola	AGO	5.2	High
Antigua and Barbuda	ATG	2.3	Low
Argentina	ARG	2.9	Low
Armenia	ARM	2.9	Low
Australia	AUS	2.4	Low
Austria	AUT	1.9	Very Low
Azerbaijan	AZE	4.8	Medium
Bahamas	BHS	2.1	Very Low
Bahrain	BHR	1.4	Very Low
Bangladesh	BGD	5.7	High



INFORM Risk Dataset

Back to Agenda

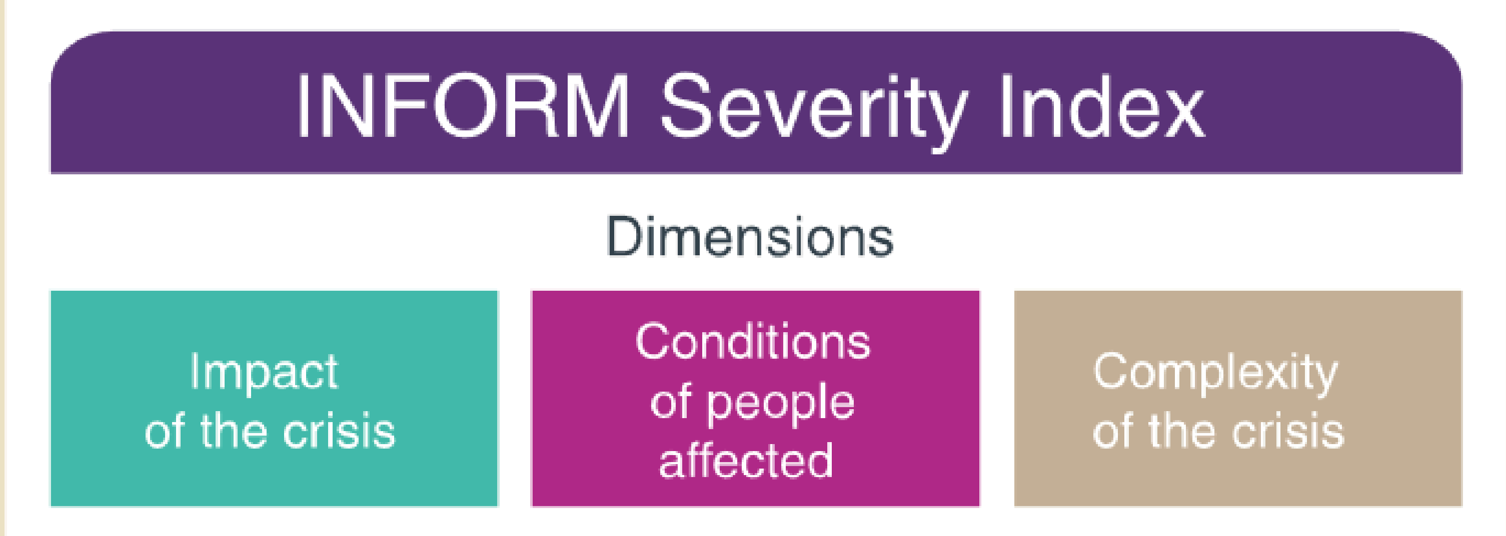
COUNTRY (a-z)	ISO3 (a-z)	INFORM RISK (0-10)	RISK CLASS (Very Low-Very High)
Afghanistan	AFG	8.1	Very High
Albania	ALB	3.1	Low
Algeria	DZA	3.6	Medium
Angola	AGO	5.2	High
Antigua and Barbuda	ATG	2.3	Low
Argentina	ARG	2.9	Low
Armenia	ARM	2.9	Low
Australia	AUS	2.4	Low
Austria	AUT	1.9	Very Low
Azerbaijan	AZE	4.8	Medium
Bahamas	BHS	2.1	Very Low
Bahrain	BHR	1.4	Very Low
Bangladesh	BGD	5.7	High



INFORM Severity Dataset

Back to Agenda

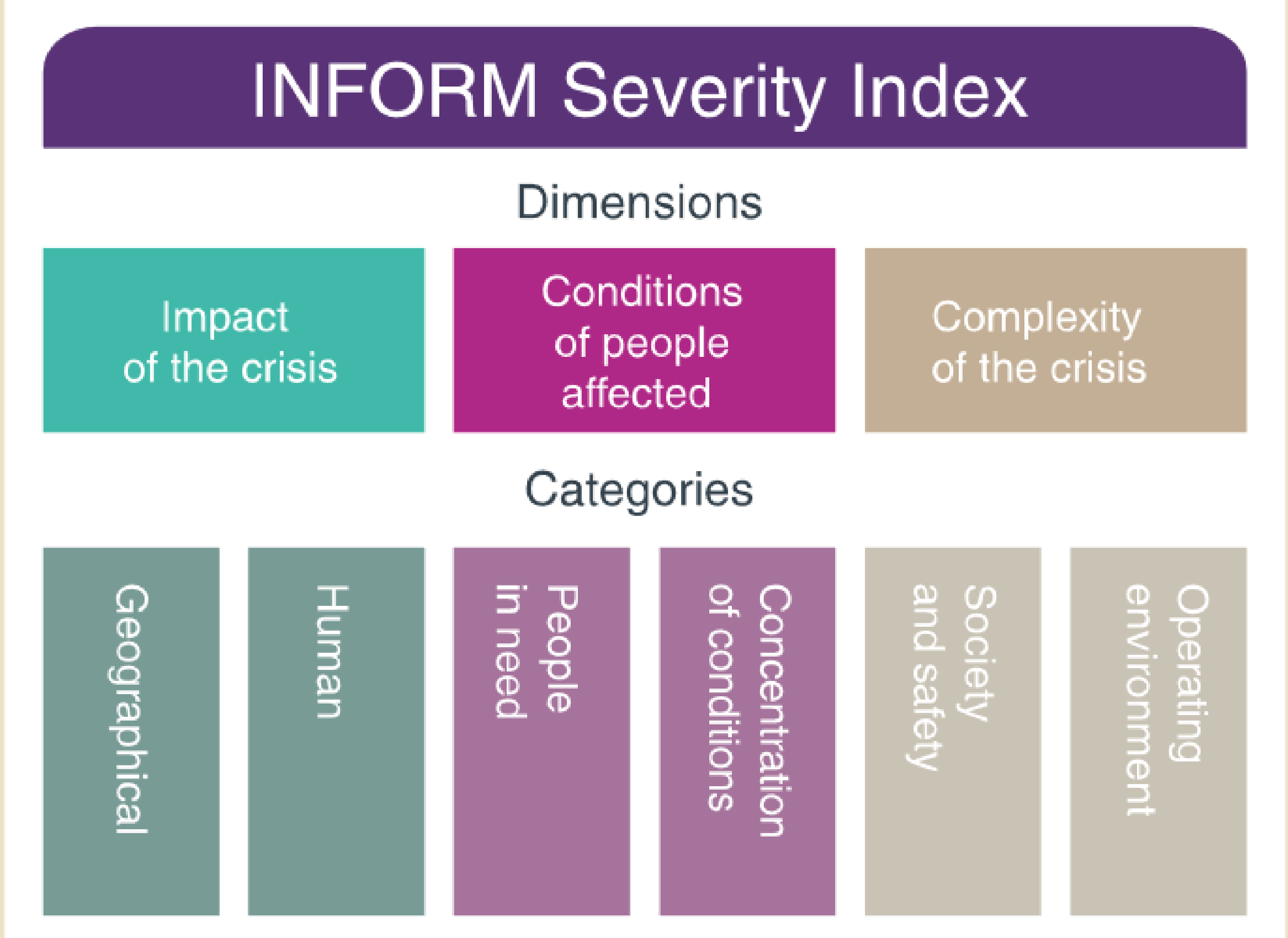
COUNTRY	CRISIS	SEVERITY (Country level)	INFORM Severity category
Afghanistan	Complex crisis in Afghanistan	4.4	Very High
Algeria	Multiple crises in Algeria	2.6	Medium
Angola	Drought in South-West Angola	3.1	High
Armenia	Nagorno-Karabakh Conflict in Armenia	2.1	Medium
Azerbaijan	Nagorno-Karabakh conflict in Azerbaijan	2.1	Medium
Bangladesh	Complex crisis in Bangladesh	4.1	Very High
Brazil	Country level Brazil	2.6	Medium
Burkina Faso	Conflict in Burkina Faso	4.0	High
Burundi	Complex in Burundi	3.3	High
Cameroon	Multiple crises in Cameroon	4.1	Very High
CAR	Complex crisis in CAR	4.2	Very High
Chad	Complex crisis in Chad	4.3	Very High



INFORM Severity Dataset

Back to Agenda

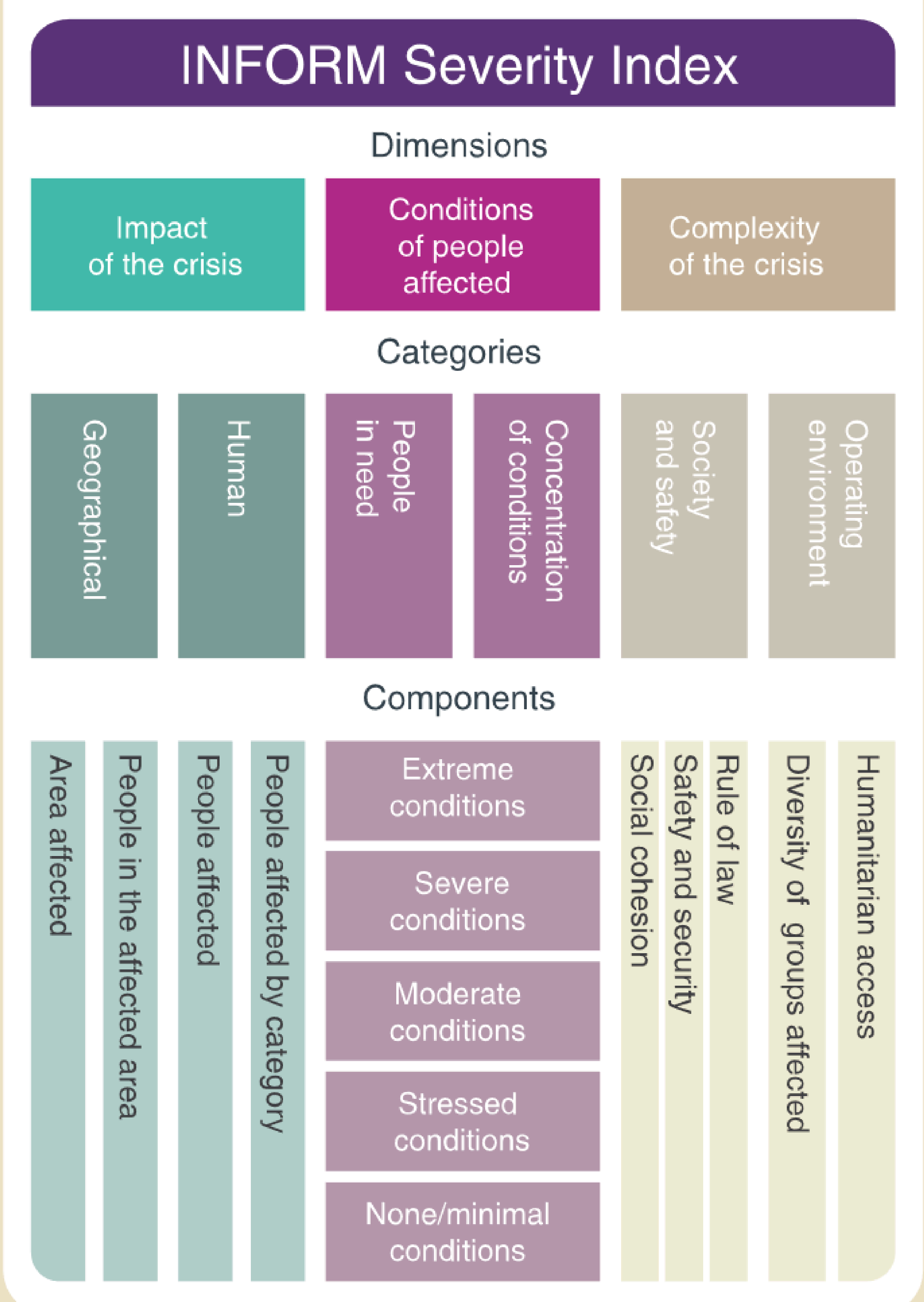
COUNTRY	CRISIS	SEVERITY (Country level)	INFORM Severity category
Afghanistan	Complex crisis in Afghanistan	4.4	Very High
Algeria	Multiple crises in Algeria	2.6	Medium
Angola	Drought in South-West Angola	3.1	High
Armenia	Nagorno-Karabakh Conflict in Armenia	2.1	Medium
Azerbaijan	Nagorno-Karabakh conflict in Azerbaijan	2.1	Medium
Bangladesh	Complex crisis in Bangladesh	4.1	Very High
Brazil	Country level Brazil	2.6	Medium
Burkina Faso	Conflict in Burkina Faso	4.0	High
Burundi	Complex in Burundi	3.3	High
Cameroon	Multiple crises in Cameroon	4.1	Very High
CAR	Complex crisis in CAR	4.2	Very High
Chad	Complex crisis in Chad	4.3	Very High



INFORM Severity Dataset

Back to Agenda

COUNTRY	CRISIS	SEVERITY (Country level)	INFORM Severity category
Afghanistan	Complex crisis in Afghanistan	4.4	Very High
Algeria	Multiple crises in Algeria	2.6	Medium
Angola	Drought in South-West Angola	3.1	High
Armenia	Nagorno-Karabakh Conflict in Armenia	2.1	Medium
Azerbaijan	Nagorno-Karabakh conflict in Azerbaijan	2.1	Medium
Bangladesh	Complex crisis in Bangladesh	4.1	Very High
Brazil	Country level Brazil	2.6	Medium
Burkina Faso	Conflict in Burkina Faso	4.0	High
Burundi	Complex in Burundi	3.3	High
Cameroon	Multiple crises in Cameroon	4.1	Very High
CAR	Complex crisis in CAR	4.2	Very High
Chad	Complex crisis in Chad	4.3	Very High



INFORM Severity Dataset

Back to Agenda

COUNTRY	CRISIS	SEVERITY (Country level)	INFORM Severity category
Afghanistan	Complex crisis in Afghanistan	4.4	Very High
Algeria	Multiple crises in Algeria	2.6	Medium
Angola	Drought in South-West Angola	3.1	High
Armenia	Nagorno-Karabakh Conflict in Armenia	2.1	Medium
Azerbaijan	Nagorno-Karabakh conflict in Azerbaijan	2.1	Medium
Bangladesh	Complex crisis in Bangladesh	4.1	Very High
Brazil	Country level Brazil	2.6	Medium
Burkina Faso	Conflict in Burkina Faso	4.0	High
Burundi	Complex in Burundi	3.3	High
Cameroon	Multiple crises in Cameroon	4.1	Very High
CAR	Complex crisis in CAR	4.2	Very High
Chad	Complex crisis in Chad	4.3	Very High

Merged Risk & Severity

Back to Agenda

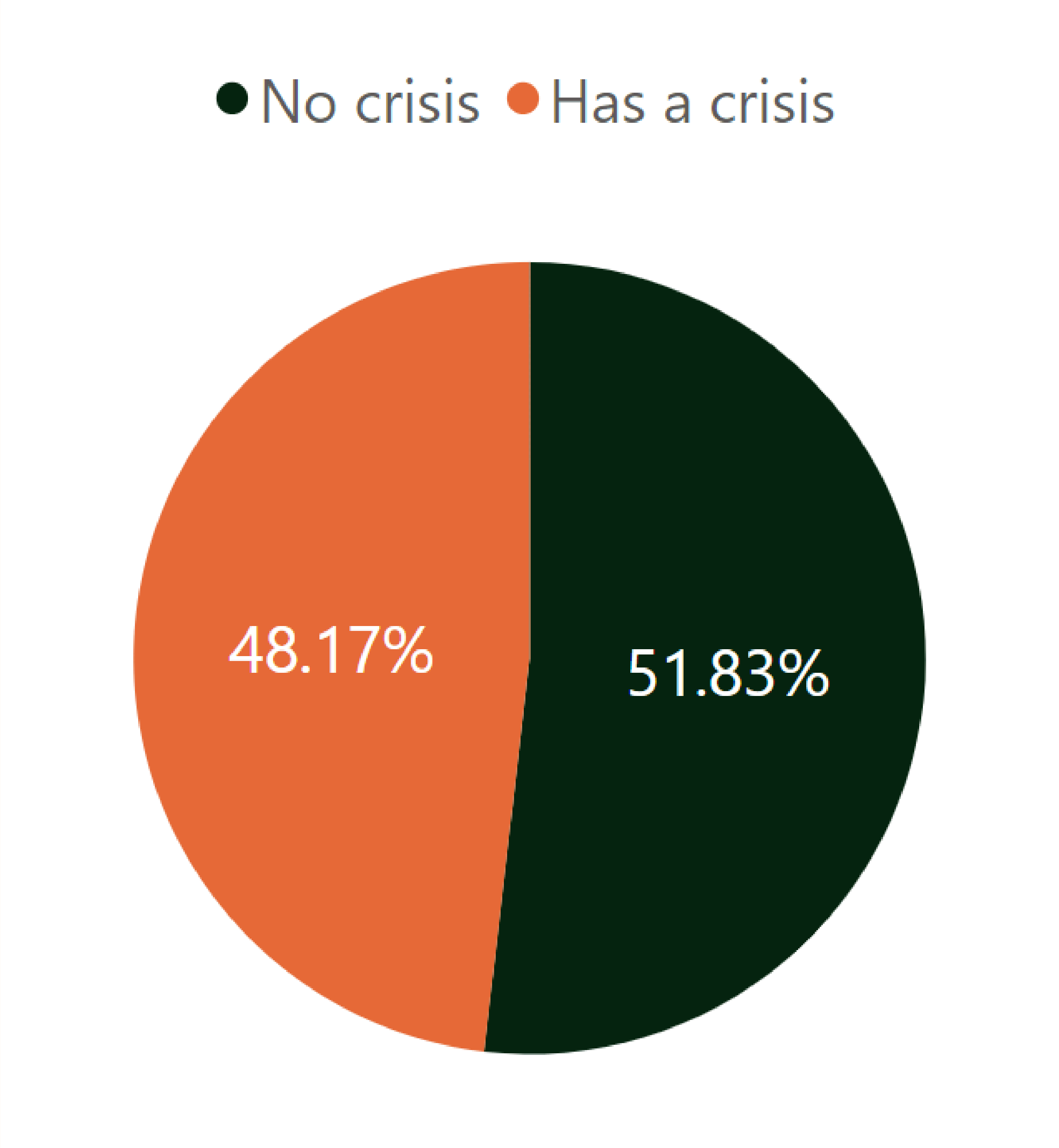
COUNTRY (a-z)	ISO3 (a-z)	INFORM RISK (0-10)	RISK CLASS (Very Low-Very High)
Afghanistan	AFG	8.1	Very High
Albania	ALB	3.1	Low
Algeria	DZA	3.6	Medium
Angola	AGO	5.2	High
Antigua and Barbuda	ATG	2.3	Low
Argentina	ARG	2.9	Low
Armenia	ARM	2.9	Low
Australia	AUS	2.4	Low
Austria	AUT	1.9	Very Low
Azerbaijan	AZE	4.8	Medium
Bahamas	BHS	2.1	Very Low
Bahrain	BHR	1.4	Very Low
Bangladesh	BGD	5.7	High

\+

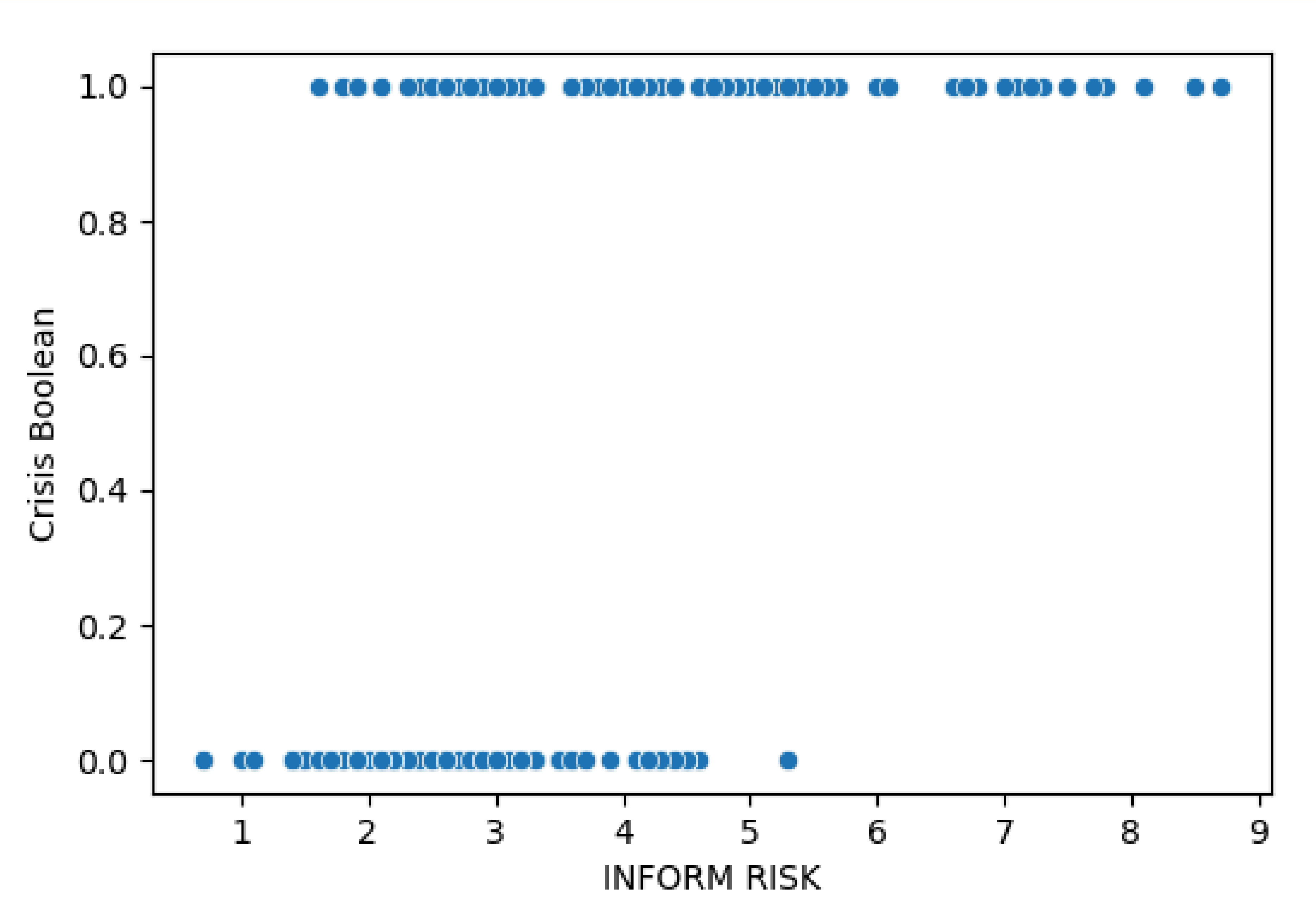
COUNTRY	CRISIS	SEVERITY (Country level)	INFORM Severity category
Afghanistan	Complex crisis in Afghanistan	4.4	Very High
Algeria	Multiple crises in Algeria	2.6	Medium
Angola	Drought in South-West Angola	3.1	High
Armenia	Nagorno-Karabakh Conflict in Armenia	2.1	Medium
Azerbaijan	Nagorno-Karabakh conflict in Azerbaijan	2.1	Medium
Bangladesh	Complex crisis in Bangladesh	4.1	Very High
Brazil	Country level Brazil	2.6	Medium
Burkina Faso	Conflict in Burkina Faso	4.0	High
Burundi	Complex in Burundi	3.3	High
Cameroon	Multiple crises in Cameroon	4.1	Very High
CAR	Complex crisis in CAR	4.2	Very High
Chad	Complex crisis in Chad	4.3	Very High

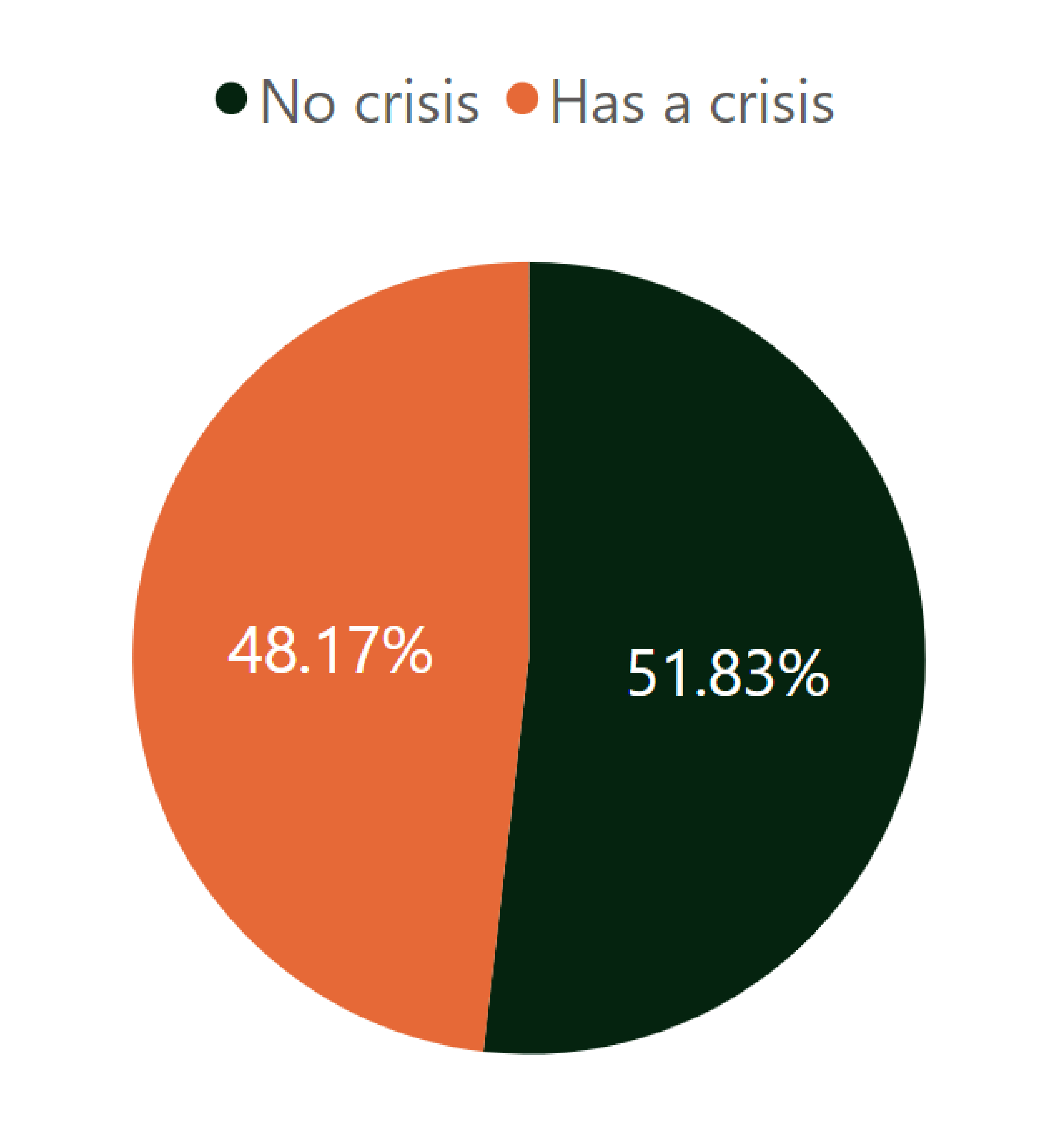
=

	COUNTRY	ISO3	INFORM RISK	RISK CLASS	Crisis Count	Crisis Boolean
0	Afghanistan	AFG	8.1	Very High	2.0	True
1	Albania	ALB	3.1	Low	0.0	False
2	Algeria	DZA	3.6	Medium	4.0	True
3	Angola	AGO	5.2	High	1.0	True
4	Antigua and Barbuda	ATG	2.3	Low	0.0	False
5	Argentina	ARG	2.9	Low	0.0	False
6	Armenia	ARM	2.9	Low	2.0	True
7	Australia	AUS	2.4	Low	0.0	False

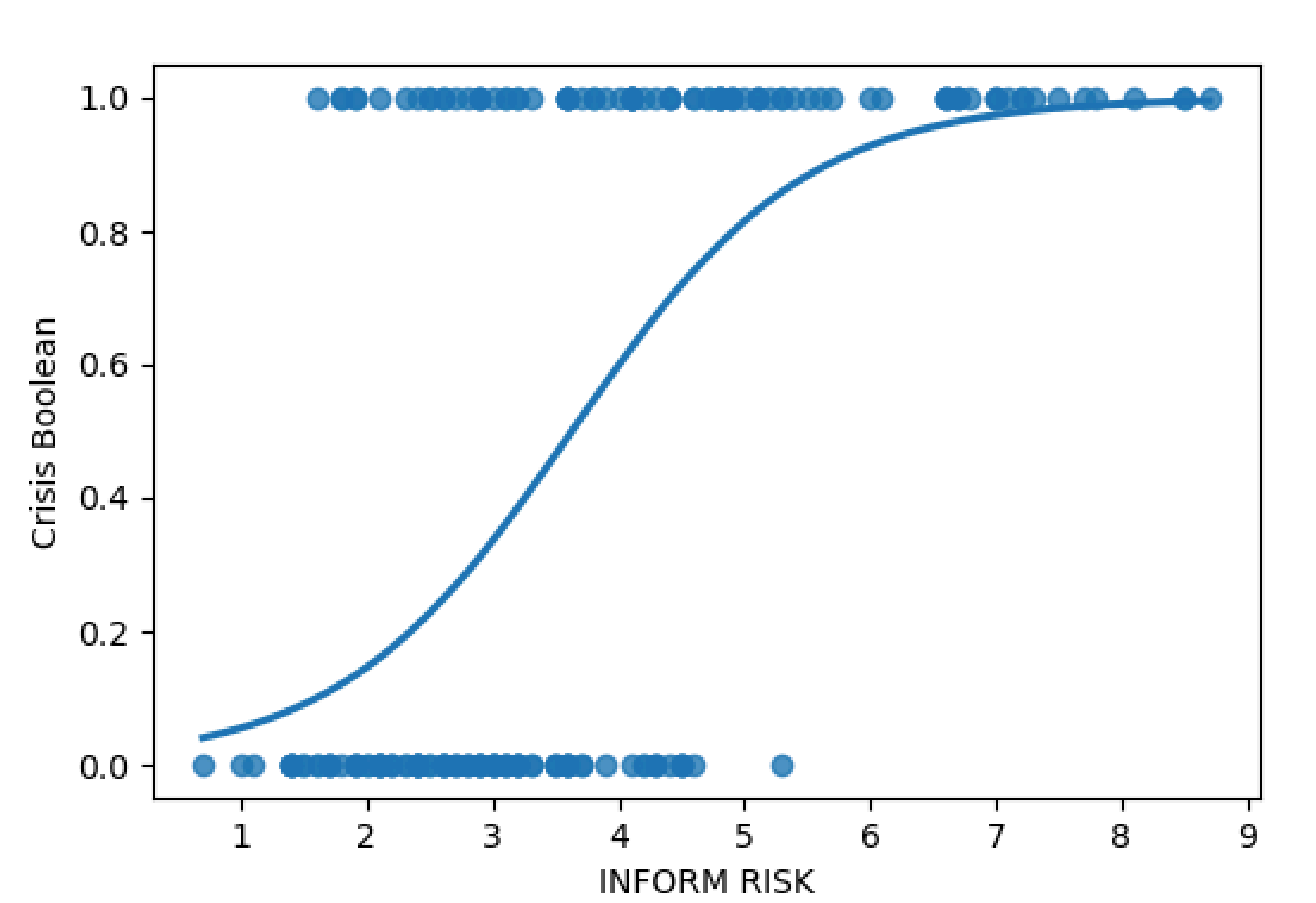


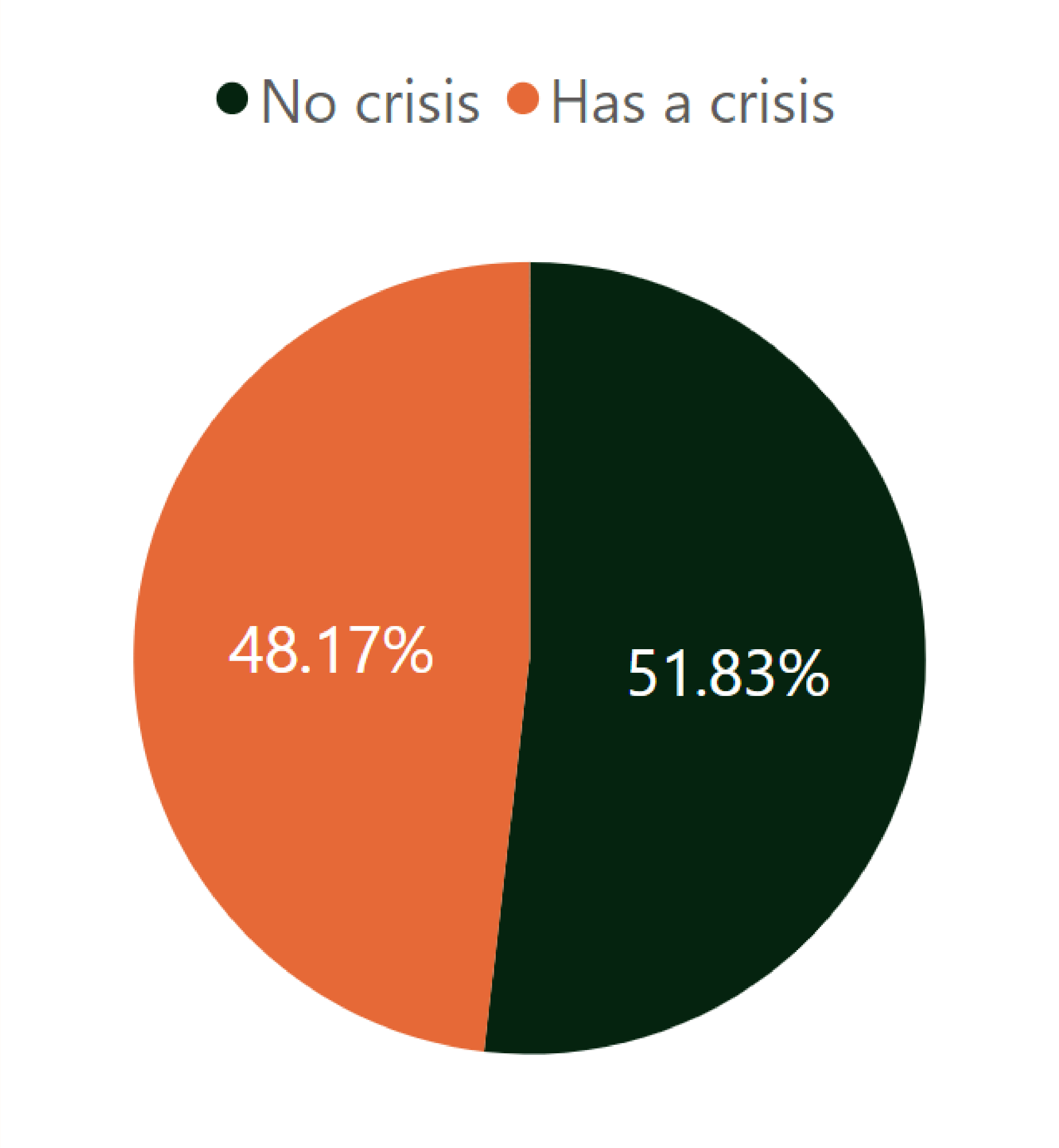
Back to Agenda



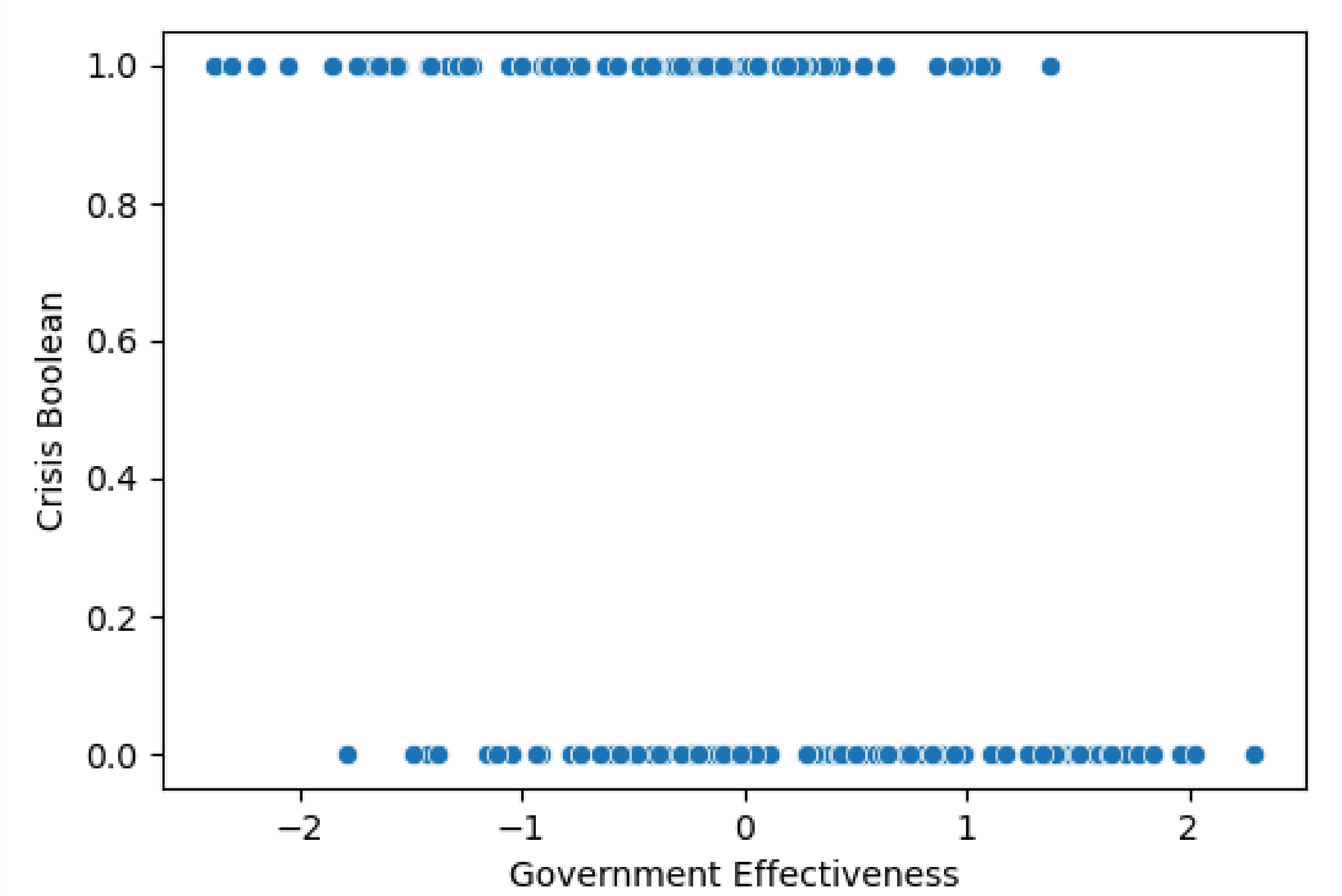


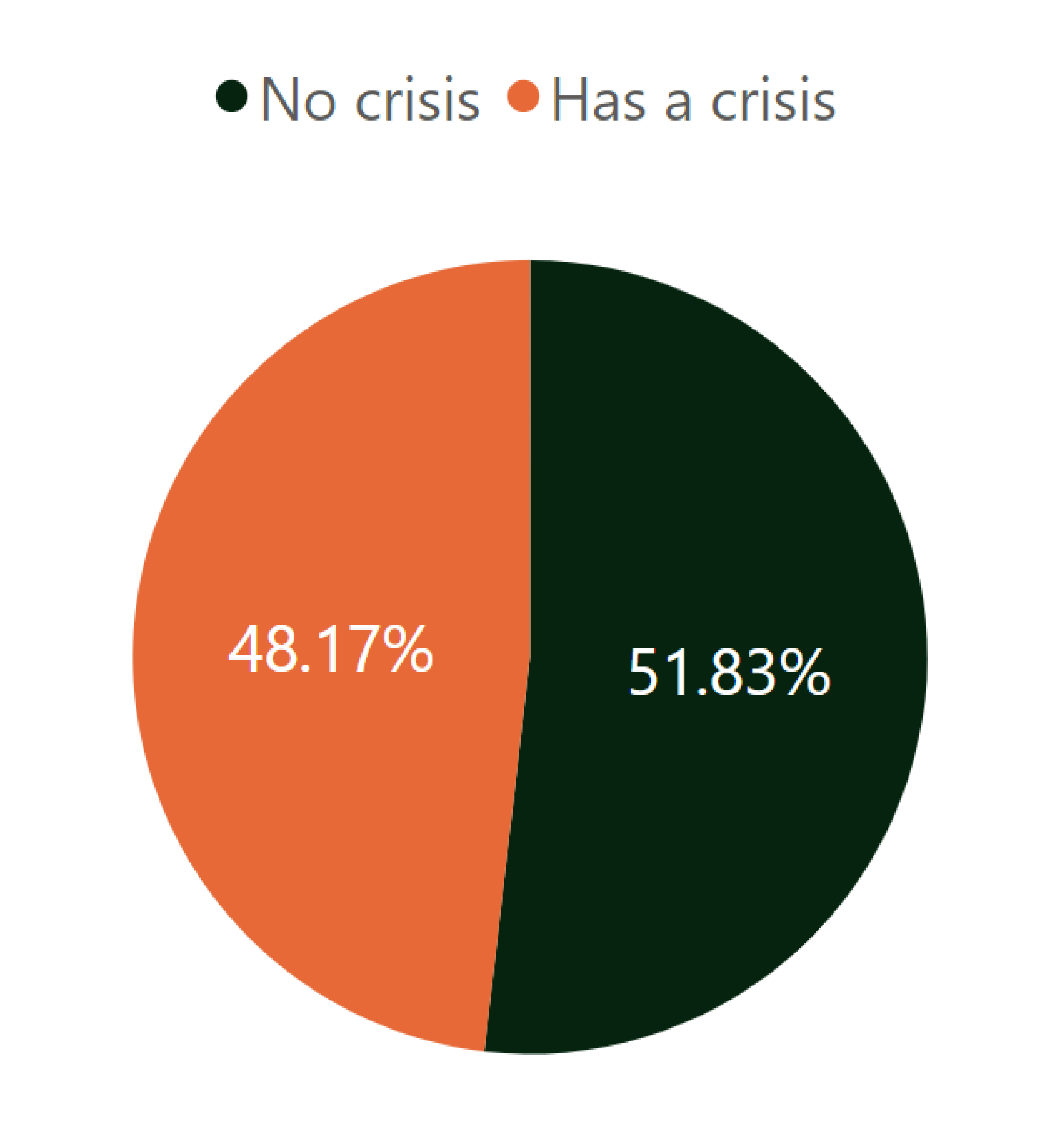
Back to Agenda



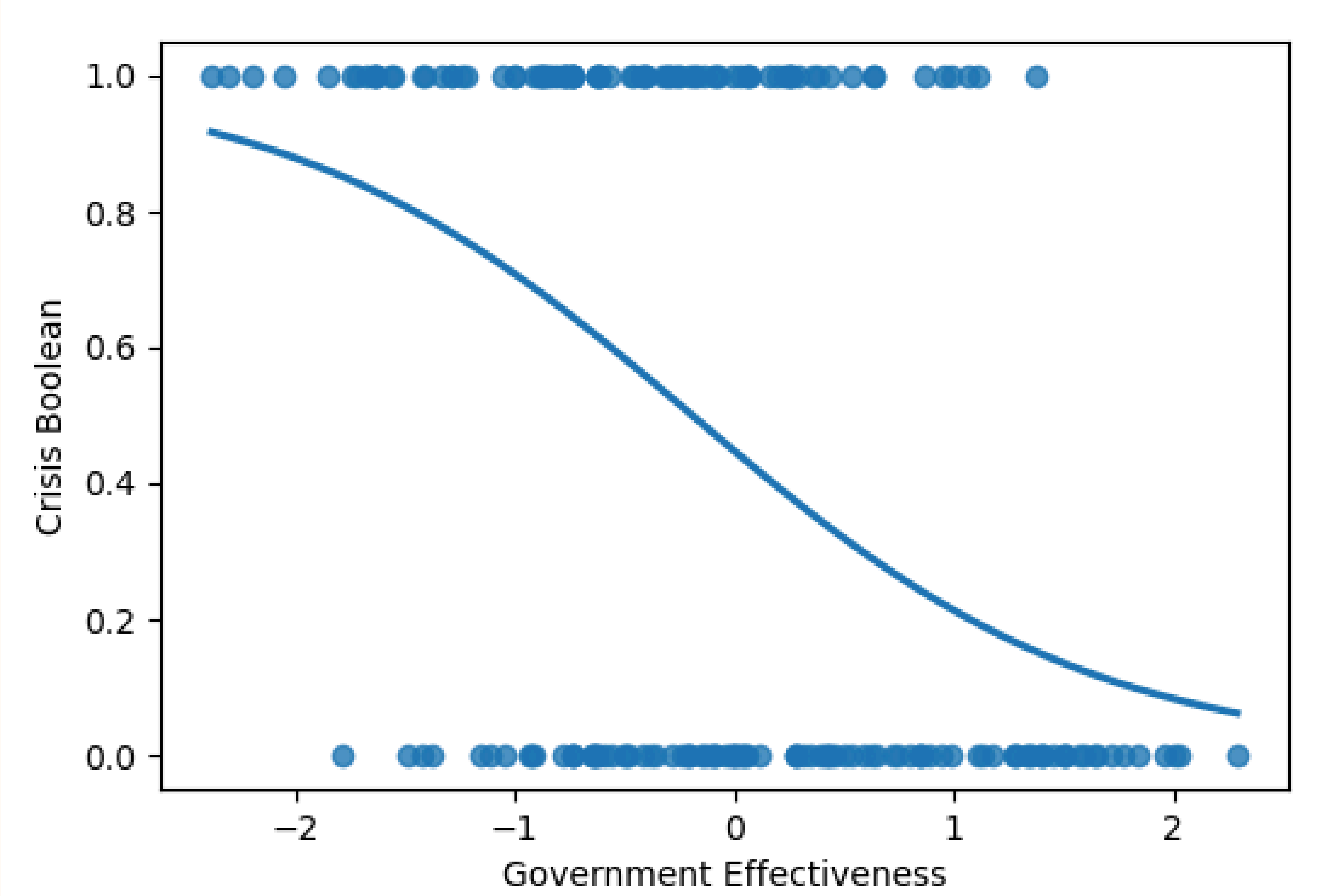


Back to Agenda

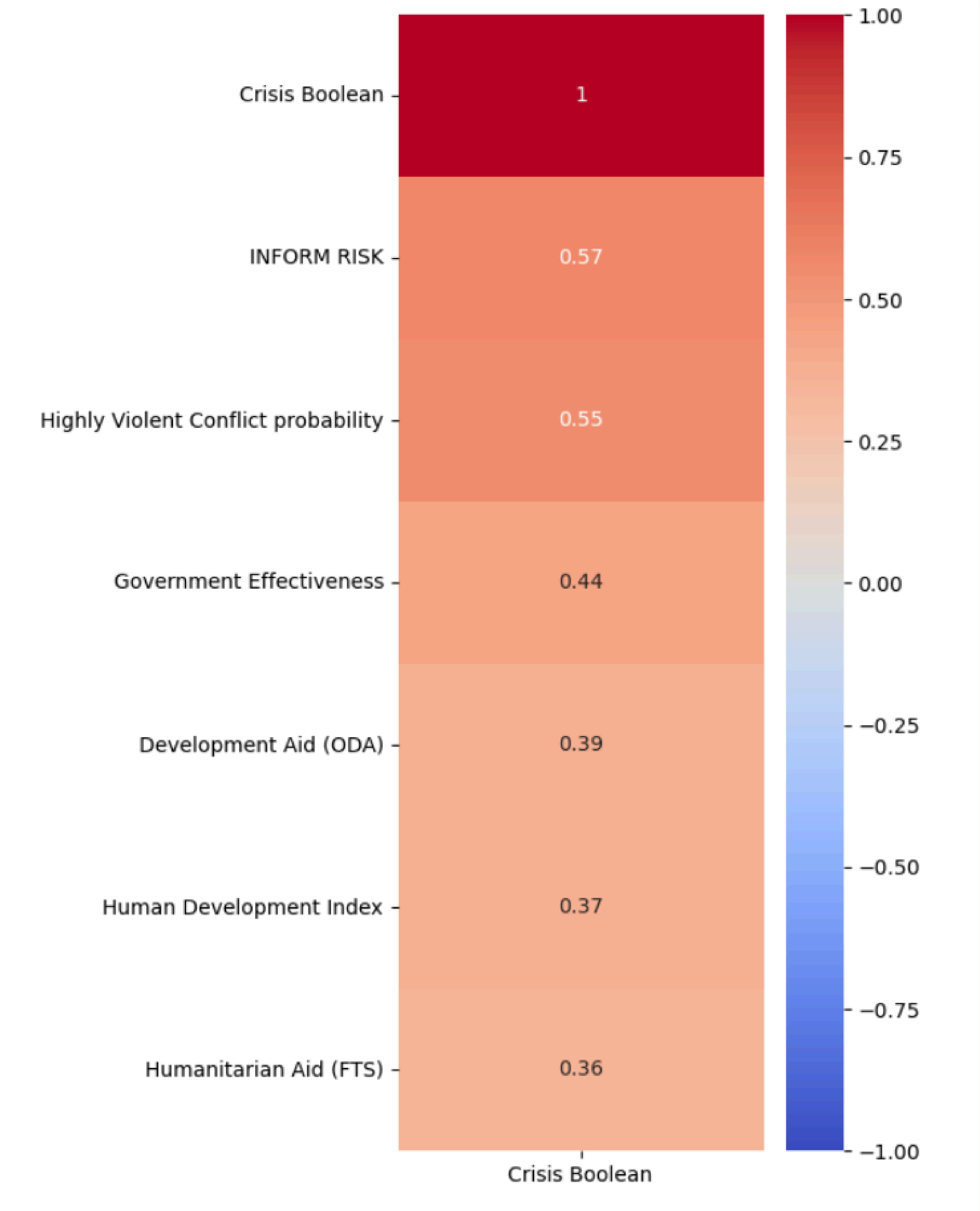




Back to Agenda



Correlations



Back to Agenda

Back to Agenda

Project 1

Explore whether the INFORM data products can be used to predict whether or not there is an ongoing humanitarian crisis

Machine Learning

ML Models

Back to Agenda

1. Logistic Regression
2. Decision Tree

Problem #1: Logistic Regression

Back to Agenda

What is logistic regression?

- Used to predict a categorical outcome, usually binary
- Fits a sigmoid curve using maximum liklihood
- Categorizes data by predicting probability of each point occuring above or below .5

Step 1: Decide what features to use in model

```
X = data[['INFORM RISK']]
y = data['Crisis Boolean']
```

```
X = data[['INFORM RISK', 'Highly Violent Conflict probability', 'Government Effectiveness',
          'Humanitarian Aid (FTS)', 'Development Aid (ODA)', 'Human Development Index']]
y = data['Crisis Boolean']
```

Step 2: Split data into training and test sets

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.2, random_state=42, stratify=y)
```

Problem #1: Logistic Regression

Back to Agenda

Step 3: Initiate the model using default parameters

```
lg = LogisticRegression()
lg.fit(X_train, y_train)
```

Step 4: Perform cross validation using KFold

What is cross validation?

- Splits data into k groups
- One group is test set, rest is training set
- Switch groups till all groups have been test set

```
k_folds = KFold(n_splits=10, random_state=1, shuffle=True)
scores = cross_val_score(lg, X, y, scoring='accuracy',cv=k_folds)
```

Problem #1: Decision Tree

Back to Agenda

What is a decision tree?

- Typically used to predict a categorical outcome
- Starts with a root node, the feature with the most importance
- Using entropy or Gini index to make decisions

Step 1: Decide what features to use in model

```
X = data[['INFORM RISK','Highly Violent Conflict probability','Government Effectiveness',
          'Humanitarian Aid (FTS)','Development Aid (ODA)', 'Human Development Index']]
y = data['Crisis Boolean']
```

Step 2: Split data into training and test sets

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.2, random_state=42)
```

Problem #1: Decision Tree

Back to Agenda

Step 3: Initiate the model using default parameters

```
dt = DecisionTreeClassifier()
dt.fit(X_train, y_train)
```

Step 4: Find the best hyperparamers using GridSearchCV hyperparameter tuning

What is GridSearchCV?

- Loops through all hyperparamers
- Output is best performing decision tree

Problem #1: Decision Tree

Back to Agenda

Step 4: Find the best hyperparamers using GridSearchCV hyperparameter tuning

```
from sklearn.model_selection import GridSearchCV
```

```
params = {
    'max_depth': [2, 3, 5, 10, 20],
    'min_samples_leaf': [5, 10, 20, 50, 100],
    'criterion': ["gini", "entropy"]
}
```

```
grid_search = GridSearchCV(estimator=dt,
                           param_grid=params,
                           cv=4, n_jobs=-1, verbose=1, scoring = "accuracy")
```

```
grid = grid_search.fit(X_train, y_train)
```

Best hyperparameters: max_depth=2, min_samples_leaf=10

Problem #1: Results

Is there an ongoing crisis?

YES

NO

Back to Agenda

Logistic Regression	Logisitic Regression with Cross Validation

With just the INFORM Risk

training accuracy score: 0.77

	precision	recall	f1-score	support
False	0.82	0.70	0.76	20
True	0.73	0.84	0.78	19
accuracy			0.77	39
macro avg	0.78	0.77	0.77	39
weighted avg	0.78	0.77	0.77	39

Multivariate

training accuracy score: 0.87

	precision	recall	f1-score	support
False	0.88	0.81	0.84	26
True	0.67	0.77	0.71	13
accuracy			0.79	39
macro avg	0.77	0.79	0.78	39
weighted avg	0.81	0.79	0.80	39

Accuracy: 0.83

Problem #1: Results

Back to Agenda

Decision Tree

training accuracy score: 1.0

	precision	recall	f1-score	support
False	0.81	0.65	0.72	26
True	0.50	0.69	0.58	13
accuracy			0.67	39
macro avg	0.65	0.67	0.65	39
weighted avg	0.71	0.67	0.68	39

Decision Tree with Hyperparameter Tuning

training accuracy score: 0.89

	precision	recall	f1-score	support
False	0.86	0.73	0.79	26
True	0.59	0.77	0.67	13
accuracy			0.74	39
macro avg	0.73	0.75	0.73	39
weighted avg	0.77	0.74	0.75	39

Decision Tree with Cross Validation

Accuracy score: 0.81

Problem #1: Results

Back to Agenda

Logistic Regression with Cross Validation	Decision Tree with Cross Validation
Accuracy: 0.83	Accuracy score: 0.81

Back to Agenda

Project 2

Explore whether the INFORM data products can be used to predict the severity of humanitarian crises

Data Cleaning & EDA

Merged Risk & Severity

Back to Agenda

COUNTRY (a-z)	ISO3 (a-z)	INFORM RISK (0-10)	RISK CLASS (Very Low-Very High)
Afghanistan	AFG	8.1	Very High
Albania	ALB	3.1	Low
Algeria	DZA	3.6	Medium
Angola	AGO	5.2	High
Antigua and Barbuda	ATG	2.3	Low
Argentina	ARG	2.9	Low
Armenia	ARM	2.9	Low
Australia	AUS	2.4	Low
Austria	AUT	1.9	Very Low
Azerbaijan	AZE	4.8	Medium
Bahamas	BHS	2.1	Very Low
Bahrain	BHR	1.4	Very Low
Bangladesh	BGD	5.7	High

+

COUNTRY	CRISIS	SEVERITY (Country level)	INFORM Severity category
Afghanistan	Complex crisis in Afghanistan	4.4	Very High
Algeria	Multiple crises in Algeria	2.6	Medium
Angola	Drought in South-West Angola	3.1	High
Armenia	Nagorno-Karabakh Conflict in Armenia	2.1	Medium
Azerbaijan	Nagorno-Karabakh conflict in Azerbaijan	2.1	Medium
Bangladesh	Complex crisis in Bangladesh	4.1	Very High
Brazil	Country level Brazil	2.6	Medium
Burkina Faso	Conflict in Burkina Faso	4.0	High
Burundi	Complex in Burundi	3.3	High
Cameroon	Multiple crises in Cameroon	4.1	Very High
CAR	Complex crisis in CAR	4.2	Very High
Chad	Complex crisis in Chad	4.3	Very High

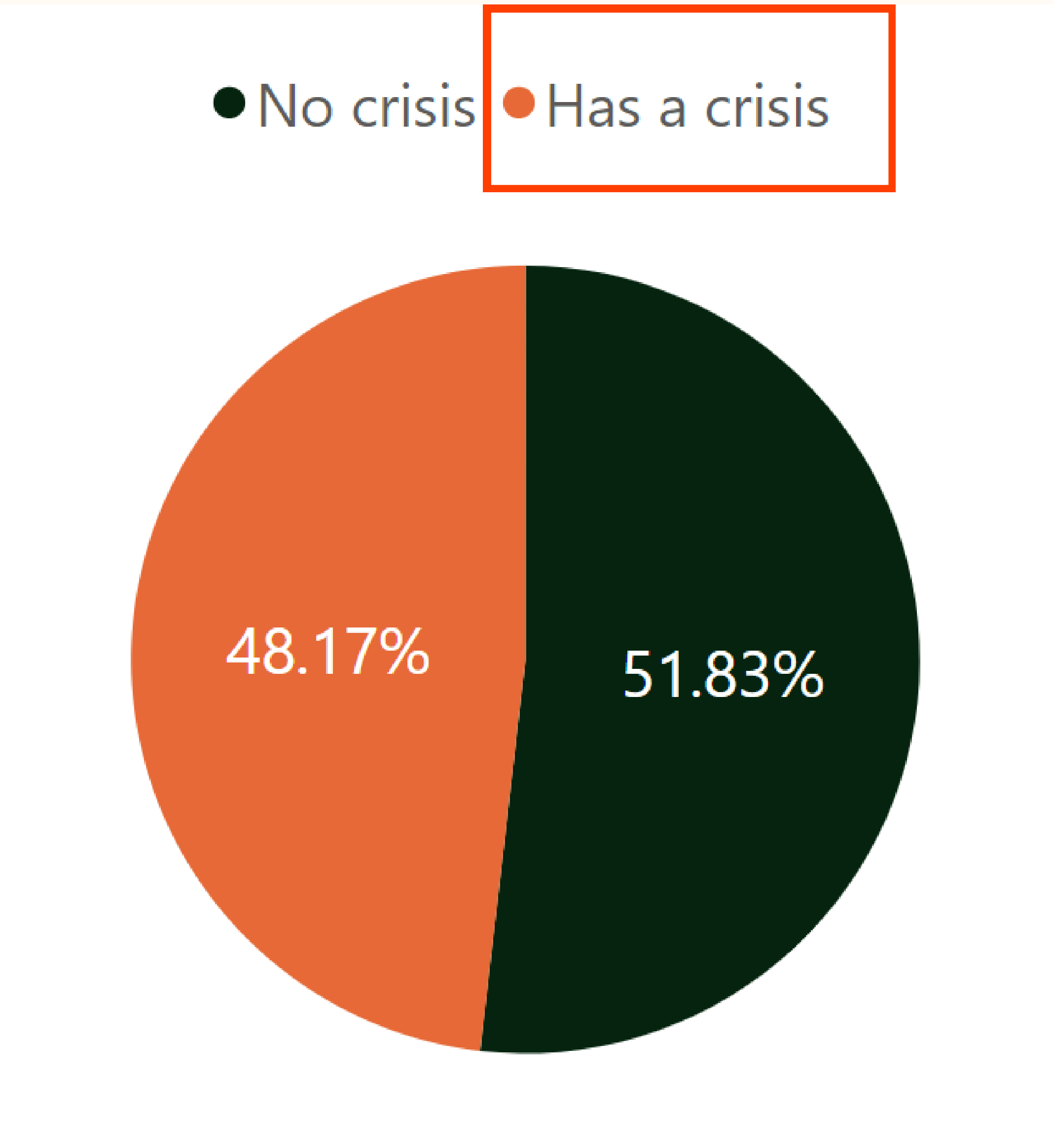
=

	COUNTRY	ISO3	INFORM RISK	RISK CLASS	Crisis Count	Crisis Boolean	Average Crisis Severity Index	Average Crisis Severity Category
0	Afghanistan	AFG	8.1	Very High	2.0	True	3.6	High
1	Albania	ALB	3.1	Low	0.0	False	NaN	NaN
2	Algeria	DZA	3.6	Medium	4.0	True	2.6	Medium
3	Angola	AGO	5.2	High	1.0	True	3.1	High
4	Antigua and Barbuda	ATG	2.3	Low	0.0	False	NaN	NaN
5	Argentina	ARG	2.9	Low	0.0	False	NaN	NaN
6	Armenia	ARM	2.9	Low	2.0	True	2.5	Medium
7	Australia	AUS	2.4	Low	0.0	False	NaN	NaN

Average Crisis Severity Index

Back to Agenda

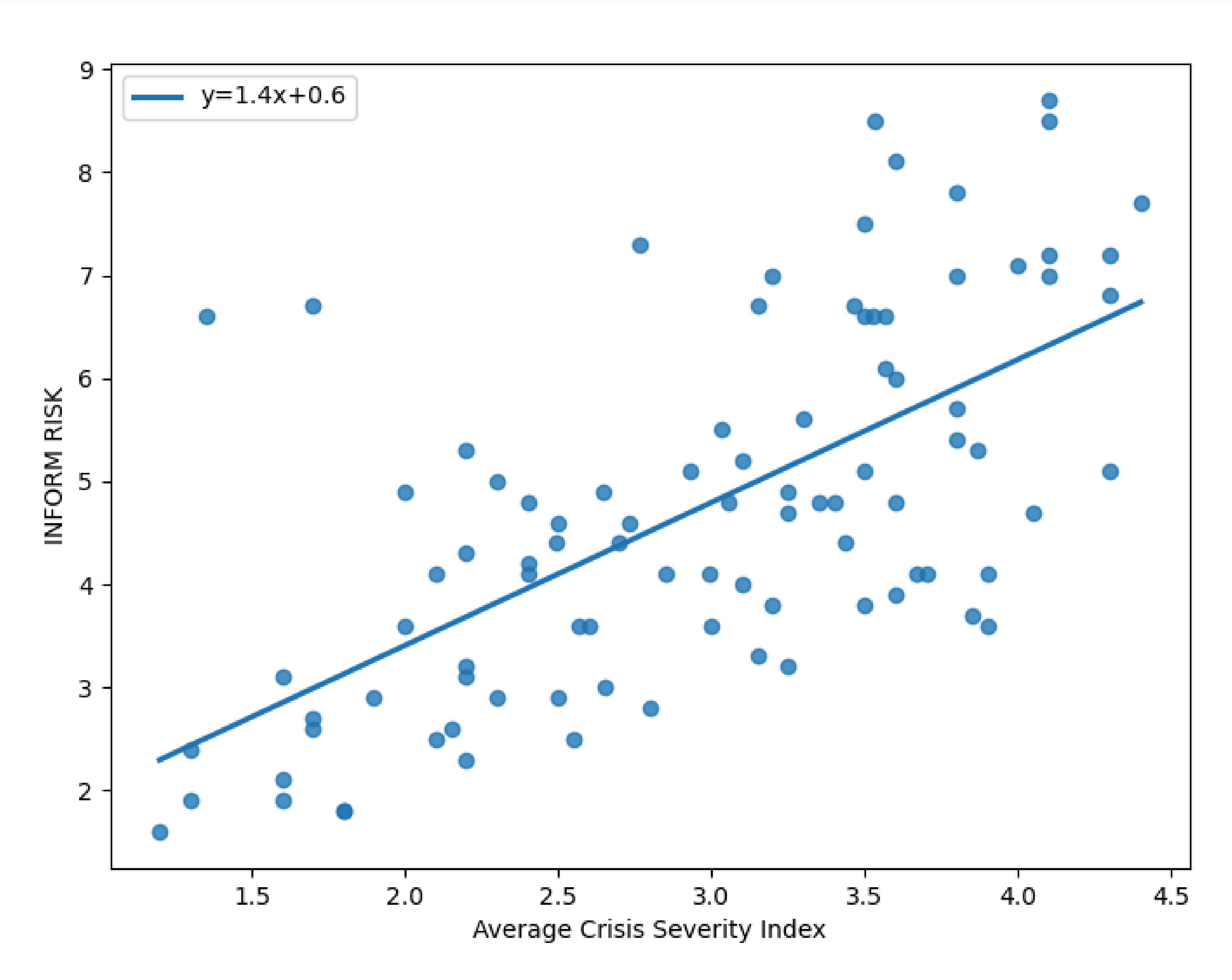
	COUNTRY	ISO3	INFORM RISK	RISK CLASS	Crisis Count	Crisis Boolean	Average Crisis Severity Index	Average Crisis Severity Category
0	Afghanistan	AFG	8.1	Very High	2.0	True	3.6	High
1	Albania	ALB	3.1	Low	0.0	False	NaN	NaN
2	Algeria	DZA	3.6	Medium	4.0	True	2.6	Medium
3	Angola	AGO	5.2	High	1.0	True	3.1	High
4	Antigua and Barbuda	ATG	2.3	Low	0.0	False	NaN	NaN
5	Argentina	ARG	2.9	Low	0.0	False	NaN	NaN
6	Armenia	ARM	2.9	Low	2.0	True	2.5	Medium
7	Australia	AUS	2.4	Low	0.0	False	NaN	NaN

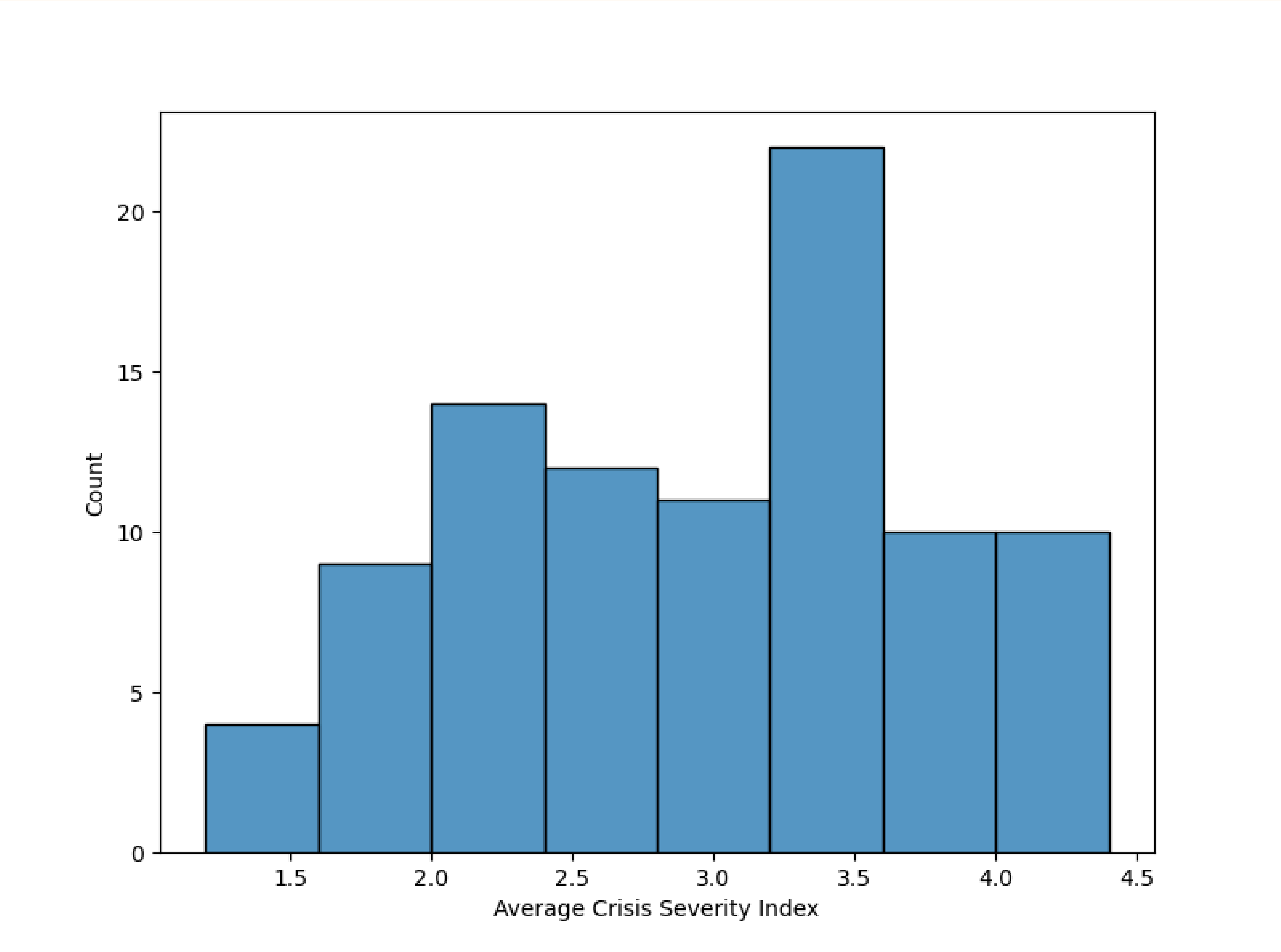


>

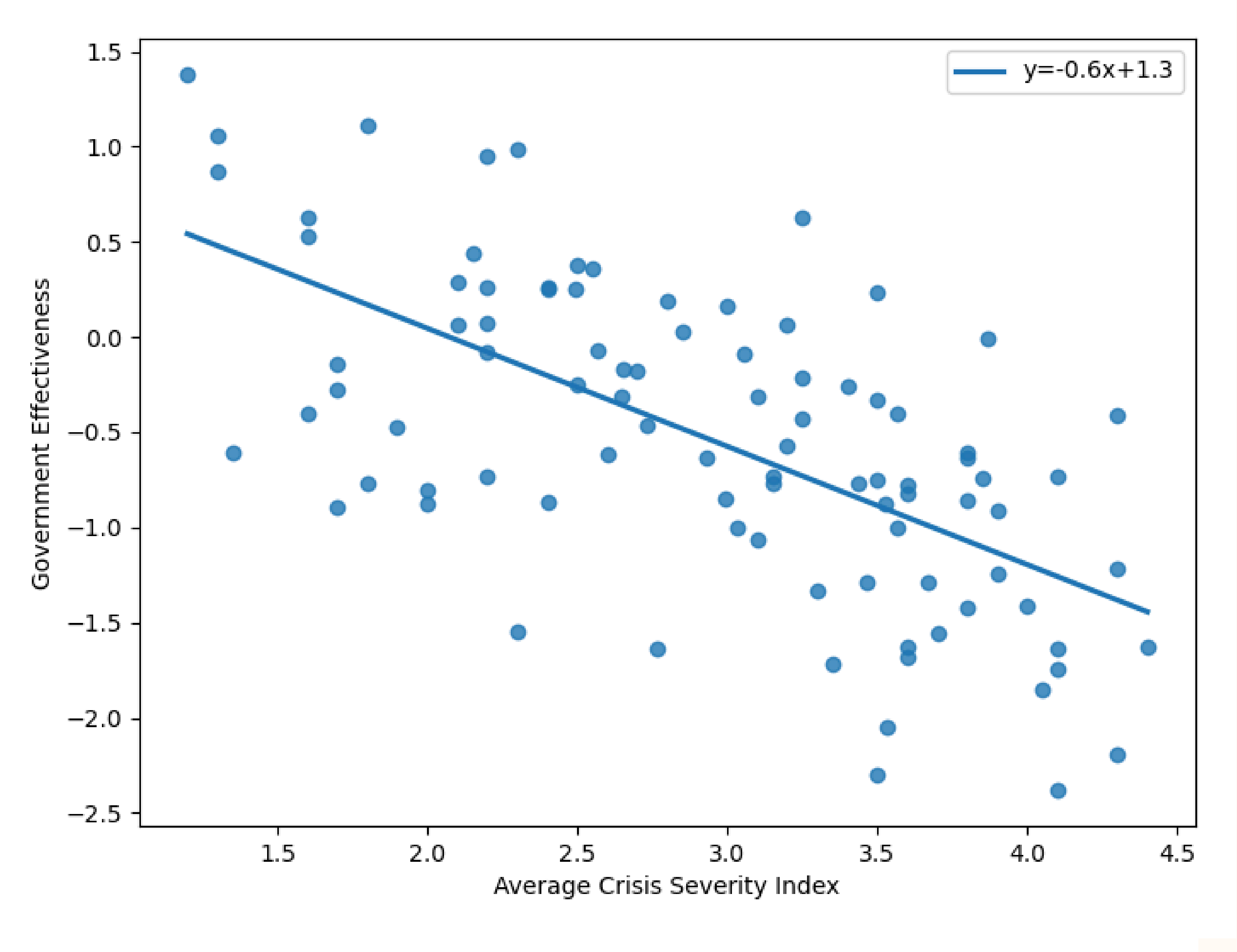
	COUNTRY	ISO3	INFORM RISK	RISK CLASS	Crisis Count	Crisis Boolean	Average Crisis Severity Index	Average Crisis Severity Category
0	Afghanistan	AFG	8.1	Very High	2.0	True	3.600000	High
1	Algeria	DZA	3.6	Medium	4.0	True	2.600000	Medium
2	Angola	AGO	5.2	High	1.0	True	3.100000	High
3	Armenia	ARM	2.9	Low	2.0	True	2.500000	Medium
4	Azerbaijan	AZE	4.8	Medium	2.0	True	2.400000	Medium
5	Bangladesh	BGD	5.7	High	3.0	True	3.800000	High
6	Belarus	BLR	1.8	Very Low	1.0	True	1.800000	Low
7	Brazil	BRA	4.6	Medium	3.0	True	2.733333	Medium

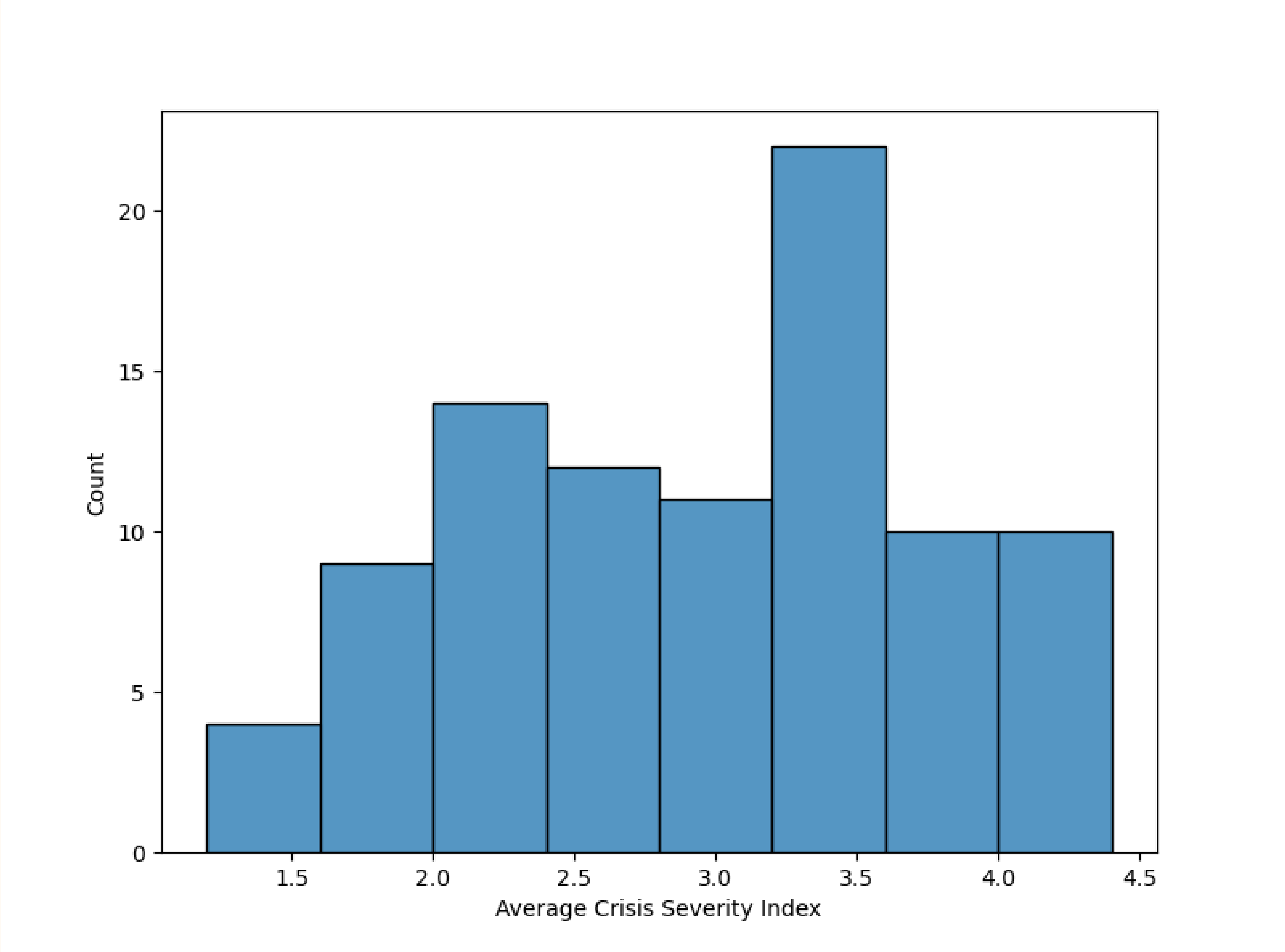
Back to Agenda



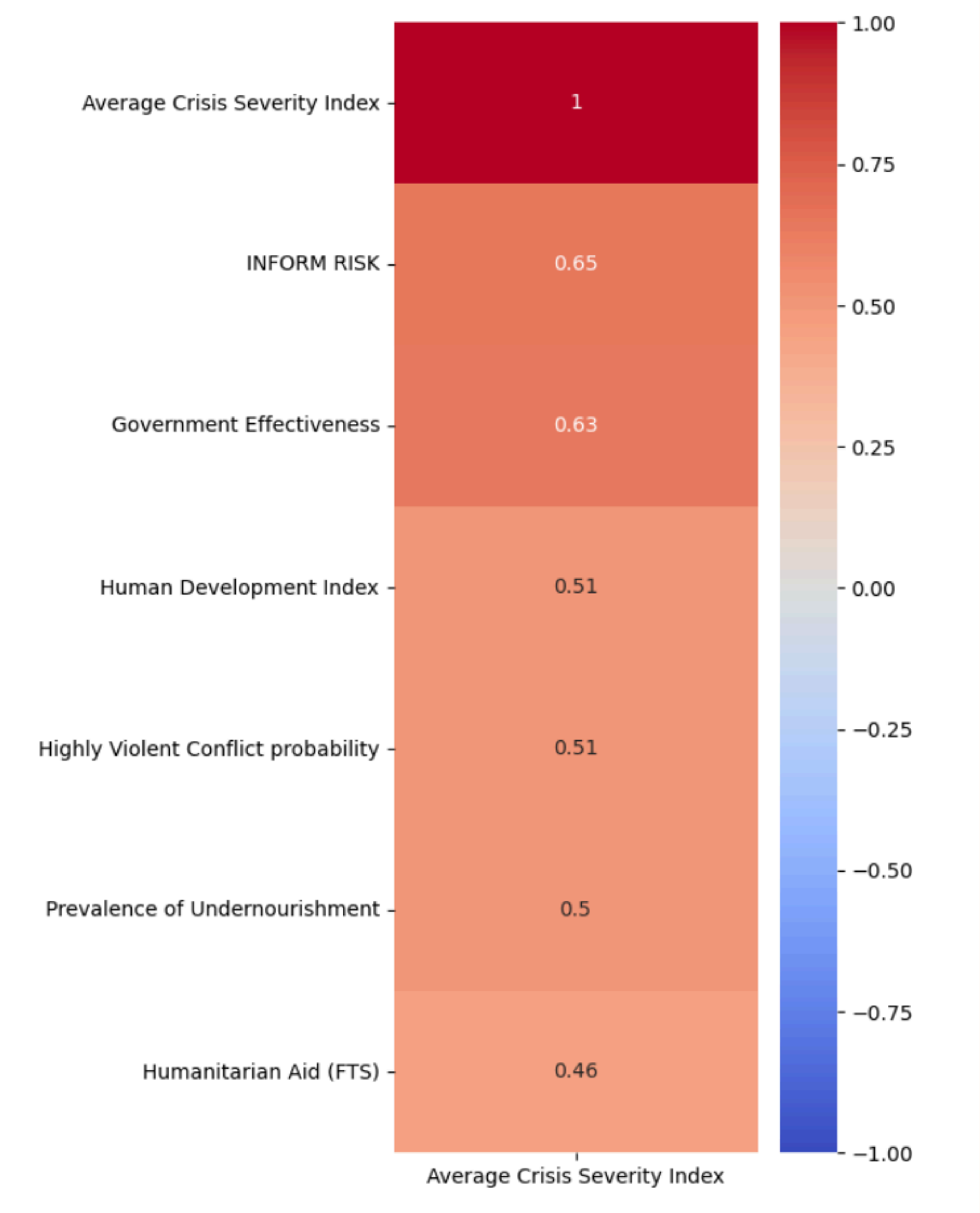


Back to Agenda





Correlations



Back to Agenda

Back to Agenda

Project 2

Explore whether the INFORM data products can be used to predict the severity of humanitarian crises

Machine Learning

ML Models

Among the countries with crises, what is the average crisis severity?

Index between 1 - 5 (least sever - most severe)

1. Linear Regression
2. XGBoost Regression

Problem #2: Linear Regression

Back to Agenda

Among the countries with crises, what is the average crisis severity?

Index between 1 - 5 (least sever - most severe)

What is linear regression?

- Used to predict a continuous outcome
- Univariate or multivariate
- Line of best fit = least sum of squared errors

Step 1: Decide what features to use in model

```
X = crisis[['INFORM RISK','Government Effectiveness', 'Human Development Index', 'Highly Violent Conflict probability',
           'Prevalence of Undernourishment','Humanitarian Aid (FTS)','Development Aid (ODA)', 'Children under 5 (% of population)',
           'Internally displaced persons (IDPs)', 'Mortality rate, under-5', 'Mobile cellular subscriptions']]
y = crisis['Average Crisis Severity Index']
```

Step 2: Split data into training and test sets

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.1, random_state=42)
```

Step 3: Initiate the model

```
lr = LinearRegression()
lr.fit(X_train, y_train)
```

Problem #2: XGBoost Regression

Back to Agenda

Among the countries with crises, what is the average crisis severity?

Index between 1 - 5 (least sever - most severe)

What is XGBoost regression?

- Can be used to predict continuous or categorical data
- Tree where root node is data average
- Calculates *residual* = the difference between observed value and predicted value
- Predict the residuals and scale them by a learning rate to reduce overfitting
- Learning rate ensures it moves with baby steps (ex: .1)
- Prediction = initial node + learning rate*residual
- Get new residual, keep adding a residual to get new predictions

```
boostR = XGBRegressor()
boostR.fit(X_train, y_train)
```

Problem #2: Results

Among the countries with crises, what is the average crisis severity?

Back to Agenda

Index between 1 - 5 (least severe - most severe)

Linear Regression	XGBoost Regression
Training R*2: **0.51**	Training R*2: **0.99**
Testing R*2: **0.52**	Testing R*2: **0.81**
MAE: **0.47**	MAE: **0.29**
RMSE: **0.54**	RMSE: **0.34**

Back to Agenda

Project 3

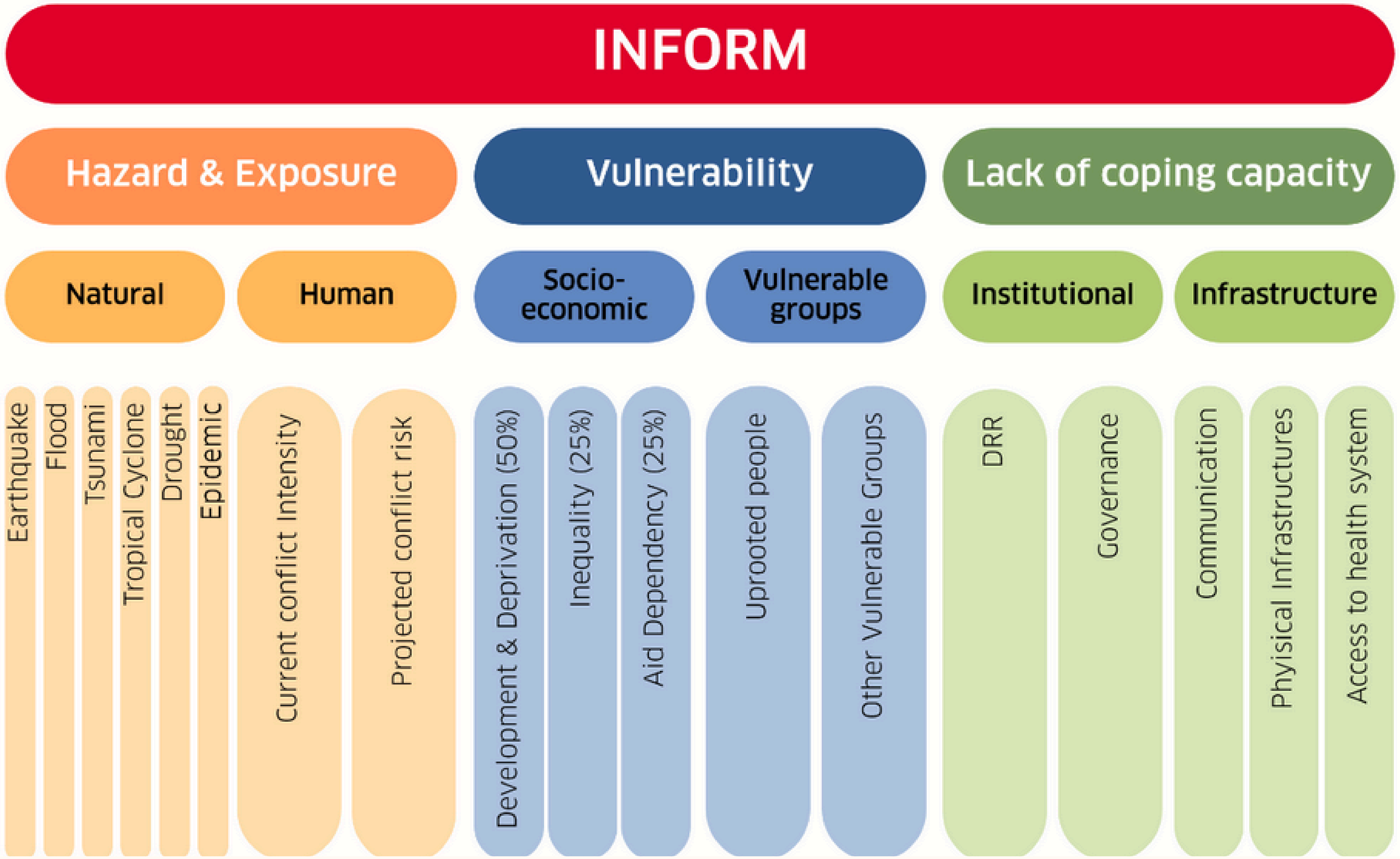
Predict & forecast how INFORM risk data changes over time

Data Cleaning & EDA

INFORM Risk Dataset

Back to Agenda

COUNTRY (a-z)	ISO3 (a-z)	INFORM RISK (0-10)	RISK CLASS (Very Low-Very High)
Afghanistan	AFG	8.1	Very High
Albania	ALB	3.1	Low
Algeria	DZA	3.6	Medium
Angola	AGO	5.2	High
Antigua and Barbuda	ATG	2.3	Low
Argentina	ARG	2.9	Low
Armenia	ARM	2.9	Low
Australia	AUS	2.4	Low
Austria	AUT	1.9	Very Low
Azerbaijan	AZE	4.8	Medium
Bahamas	BHS	2.1	Very Low
Bahrain	BHR	1.4	Very Low
Bangladesh	BGD	5.7	High



Problem #3

Back to Agenda

Can we forecast the INFORM Risk index?

Database of INFORM Risk Indices Over Time

- Dataframe structure
- Merging Indices with Core Indicators
- Filling NaNs
- Exploring associations
- Forecasting predictions

INFORM Risk Dataset Trend 2015-2024

Back to Agenda

Iso3	IndicatorId	IndicatorName	IndicatorScore	SurveyYear	Indicator Type	INFORMYear
AFG	AFF_DR	People affected by dı	886000	2022	Core Indicators	2024
AGO	AFF_DR	People affected by dı	197920.4571	2022	Core Indicators	2024
ALB	AFF_DR	People affected by dı	91428.57143	2022	Core Indicators	2024
ARE	AFF_DR	People affected by dı	0	2022	Core Indicators	2024
ARG	AFF_DR	People affected by dı	1000.914286	2022	Core Indicators	2024
ARM	AFF_DR	People affected by dı	8485.714286	2022	Core Indicators	2024
ATG	AFF_DR	People affected by dı	0	2022	Core Indicators	2024
AUS	AFF_DR	People affected by dı	200000	2022	Core Indicators	2024
AUT	AFF_DR	People affected by dı	0	2022	Core Indicators	2024
AZE	AFF_DR	People affected by dı	0	2022	Core Indicators	2024
BDI	AFF_DR	People affected by dı	87500	2022	Core Indicators	2024
BEL	AFF_DR	People affected by dı	0	2022	Core Indicators	2024
BEN	AFF_DR	People affected by dı	0	2022	Core Indicators	2024
BFA	AFF_DR	People affected by dı	461349.0857	2022	Core Indicators	2024
BGD	AFF_DR	People affected by dı	142857.1429	2022	Core Indicators	2024
BGR	AFF_DR	People affected by dı	0	2022	Core Indicators	2024
BHR	AFF_DR	People affected by dı	0	2022	Core Indicators	2024
BHS	AFF_DR	People affected by dı	0	2022	Core Indicators	2024
BIH	AFF_DR	People affected by dı	1787.857143	2022	Core Indicators	2024
BLR	AFF_DR	People affected by dı	0	2022	Core Indicators	2024
BLZ	AFF_DR	People affected by dı	0	2022	Core Indicators	2024
EST	CC.INF.AHC.HEAL	Health expenditure p	0	2021	INORM Index	2024
ETH	CC.INF.AHC.HEAL	Health expenditure p	9.9	2020	INORM Index	2024
FIN	CC.INF.AHC.HEAL	Health expenditure p	0	2020	INORM Index	2024
FJI	CC.INF.AHC.HEAL	Health expenditure p	8.7	2020	INORM Index	2024
FRA	CC.INF.AHC.HEAL	Health expenditure p	0	2020	INORM Index	2024
FSM	CC.INF.AHC.HEAL	Health expenditure p	8.9	2020	INORM Index	2024

Split dataframe: INFORM Risk Indices & Core Indicators

```
DataFrame for Indicator Types Core Indicators:
<class 'pandas.core.frame.DataFrame'>
Index: 147420 entries, 0 to 473667
Data columns (total 7 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   Iso3            147420 non-null  object
 1   IndicatorId     147420 non-null  object
 2   IndicatorName   147420 non-null  object
 3   IndicatorScore  147420 non-null  float64
 4   SurveyYear      147420 non-null  int64
 5   Indicator Type  147420 non-null  object
 6   INFORMYear      147420 non-null  int64
dtypes: float64(1), int64(2), object(4)
memory usage: 9.0+ MB
None
```

```
DataFrame for Indicator Types INORM Index:
<class 'pandas.core.frame.DataFrame'>
Index: 326248 entries, 21 to 473143
Data columns (total 7 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   Iso3            326248 non-null  object
 1   IndicatorId     326248 non-null  object
 2   IndicatorName   325091 non-null  object
 3   IndicatorScore  326248 non-null  float64
 4   SurveyYear      326248 non-null  int64
 5   Indicator Type  326248 non-null  object
 6   INFORMYear      326248 non-null  int64
dtypes: float64(1), int64(2), object(4)
memory usage: 19.9+ MB
None
```

Melting & Pivoting 2015-2024 Trend Data

Back to Agenda

	Iso3	IndicatorName	IndicatorScore	INFORMYear
0	AFG	People affected by drought (absolute) - raw	886000.000000	2024
1	AGO	People affected by drought (absolute) - raw	197920.457143	2024
2	ALB	People affected by drought (absolute) - raw	91428.571429	2024
3	ARE	People affected by drought (absolute) - raw	0.000000	2024
4	ARG	People affected by drought (absolute) - raw	1000.914286	2024
...	...	...	...	...
473663	GNB	Population living in slums (% of urban populat...	60.810370	2024
473664	GTM	Population living in slums (% of urban populat...	37.600000	2024
473665	GUY	Population living in slums (% of urban populat...	12.140200	2024
473666	HTI	Population living in slums (% of urban populat...	48.900000	2024
473667	IDN	Population living in slums (% of urban populat...	19.410830	2024

473668 rows × 4 columns

Melt: Unpivot a DataFrame from wide to long format, optionally leaving identifiers set

```
df1 = pd.melt(df,
              id_vars=['INFORMYear', 'Iso3', 'IndicatorName'],
              var_name='IndicatorScore',
              value_name='Score')
```

```
# Pivot the melted DataFrame to create individual columns for each indicator
df1 = df1.pivot_table(index=['INFORMYear', 'Iso3'],
                     columns='IndicatorName',
                     values='Score').reset_index()
```

Dataframe Structure for Time Series Analysis

	INFORMYear	Iso3	IndicatorName	IndicatorScore	Score
0	2024	AFG	People affected by drought (absolute) - raw	IndicatorScore	886000.000000
1	2024	AGO	People affected by drought (absolute) - raw	IndicatorScore	197920.457143
2	2024	ALB	People affected by drought (absolute) - raw	IndicatorScore	91428.571429
3	2024	ARE	People affected by drought (absolute) - raw	IndicatorScore	0.000000
4	2024	ARG	People affected by drought (absolute) - raw	IndicatorScore	1000.914286
...	...	...	...	...	...
473663	2024	GNB	Population living in slums (% of urban populat...	IndicatorScore	60.810370
473664	2024	GTM	Population living in slums (% of urban populat...	IndicatorScore	37.600000
473665	2024	GUY	Population living in slums (% of urban populat...	IndicatorScore	12.140200
473666	2024	HTI	Population living in slums (% of urban populat...	IndicatorScore	48.900000
473667	2024	IDN	Population living in slums (% of urban populat...	IndicatorScore	19.410830

473668 rows × 5 columns

IndicatorName	INFORMYear	Iso3	% of Populations at risk of Plasmodium falciparum malaria	% of Populations at risk of Plasmodium vivax malaria	Access to Health Care	Access to electricity	Access to improved sanitation facilities	Access to improved water source	Adult literacy rate	Adult prevalence of HIV-AIDS	...	Urban population growth	Urban population growth (annual %)	Vets per capita
0	2015	AFG	0.872426	0.908841	8.9	45.300000	6.5	8.3	9.9	0.15	...	9.1	4.531768	0.008513
1	2015	AGO	0.996012	0.996146	7.7	19.400000	6.0	9.3	6.2	2.85	...	9.7	4.853340	0.000933
2	2015	ALB	0.000000	0.000000	3.7	49.974998	0.3	1.4	0.6	0.15	...	3.3	1.646116	0.043124
3	2015	ARE	0.000000	0.000000	1.5	50.000000	0.2	0.0	2.1	0.15	...	2.6	1.318987	0.003350
4	2015	ARG	0.000000	0.000000	1.3	50.000000	0.6	0.2	0.2	0.60	...	2.5	1.239639	0.024787
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1905	2024	WSM	0.000000	0.000000	6.3	49.250002	0.4	1.6	0.2	NaN	...	1.9	0.928355	0.004769
1906	2024	YEM	0.685021	0.419070	6.5	38.688599	5.1	7.9	NaN	0.15	...	7.6	3.796455	0.001715
1907	2024	ZAF	0.125373	0.249348	4.5	45.200002	2.4	1.2	1.1	14.15	...	3.1	1.557749	0.003141
1908	2024	ZMB	1.000000	1.000000	5.7	25.992558	7.6	6.9	2.6	10.40	...	8.0	4.009238	0.009208
1909	2024	ZWE	0.000000	0.000000	6.6	27.039964	7.2	7.5	2.1	10.80	...	4.6	2.308437	0.033776

1910 rows × 240 columns

Data Cleaning

Back to Agenda

Merge Index & Core Indicator dataframes

Create Mean INFORM Risk Index & INFORM Risk Categories Columns

Manage NaNs for Core Indicators

```
Access to Health Care' has 10 null values
Adult literacy rate' has 354 null values
Adult prevalence of HIV-AIDS' has 420 null values
Corruption Perception Index' has 137 null values
GDP per capita' has 27 null values
Gender Inequality Index' has 216 null values
Health expenditure per capita' has 44 null values
Inequality' has 103 null values
Internet users' has 8 null values
Literacy rate, adult total' has 354 null values
Maternal Mortality Ratio' has 70 null values
Multidimensional Poverty Index' has 753 null values
Palestinian Refugees' has 1870 null values
Personal remittances, received (% of GDP)' has 93 null values
Physicians density' has 27 null values
Population living in slums (% of urban population)' has 894 null values
Under-five mortality rate' has 10 null values
Vets per capita' has 85 null values
Volume of remittances' has 93 null values
```

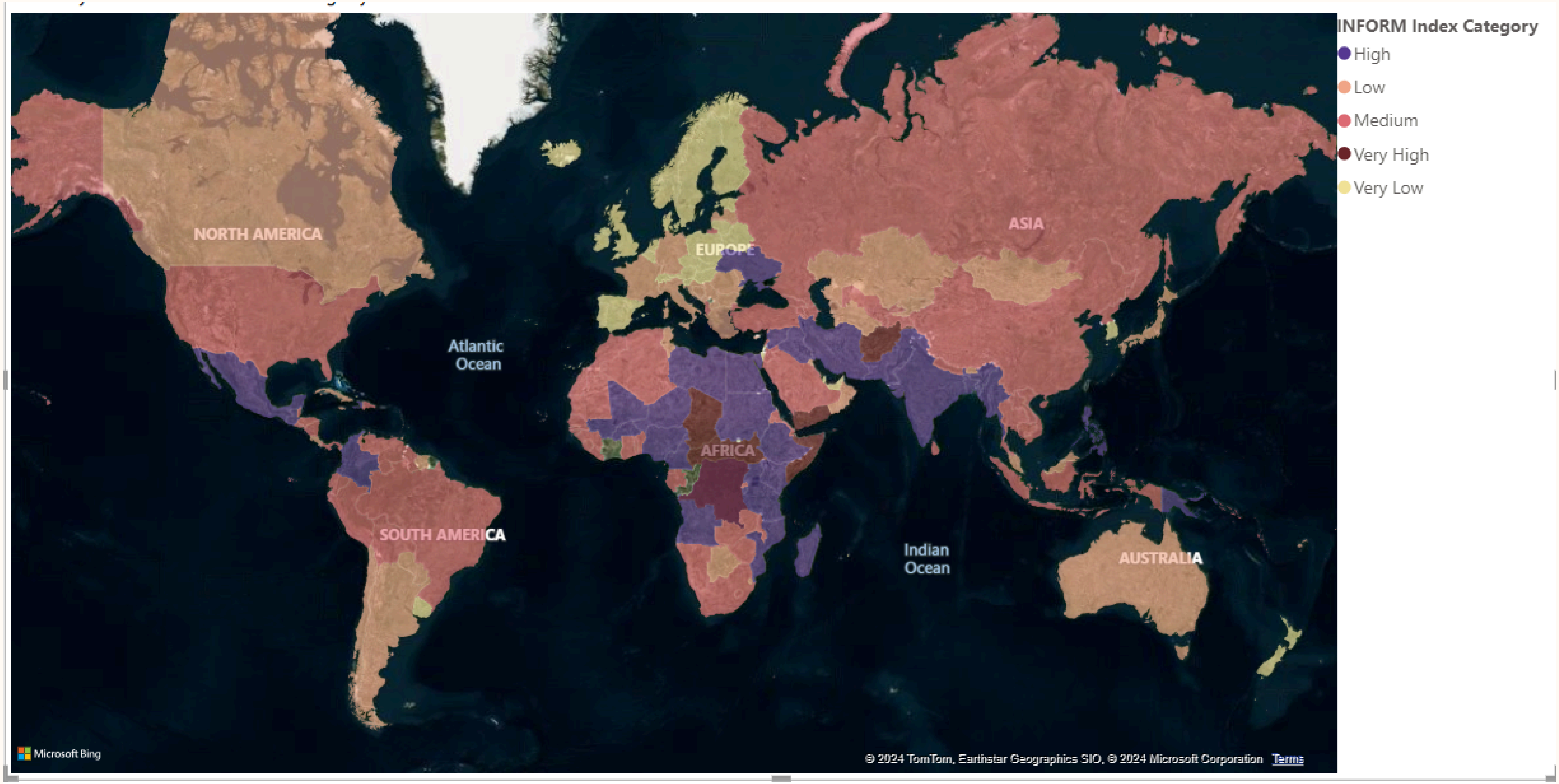
Fill as most recent values per country, or column mean where none available

For Time Series - Set Date as Index for dataframe

```
df = pd.read_csv("2015-2024_year_df_INFORM_index_core_indicators_mean_index.csv", index_col = 'INFORMYear',
                 parse_dates=True)
```

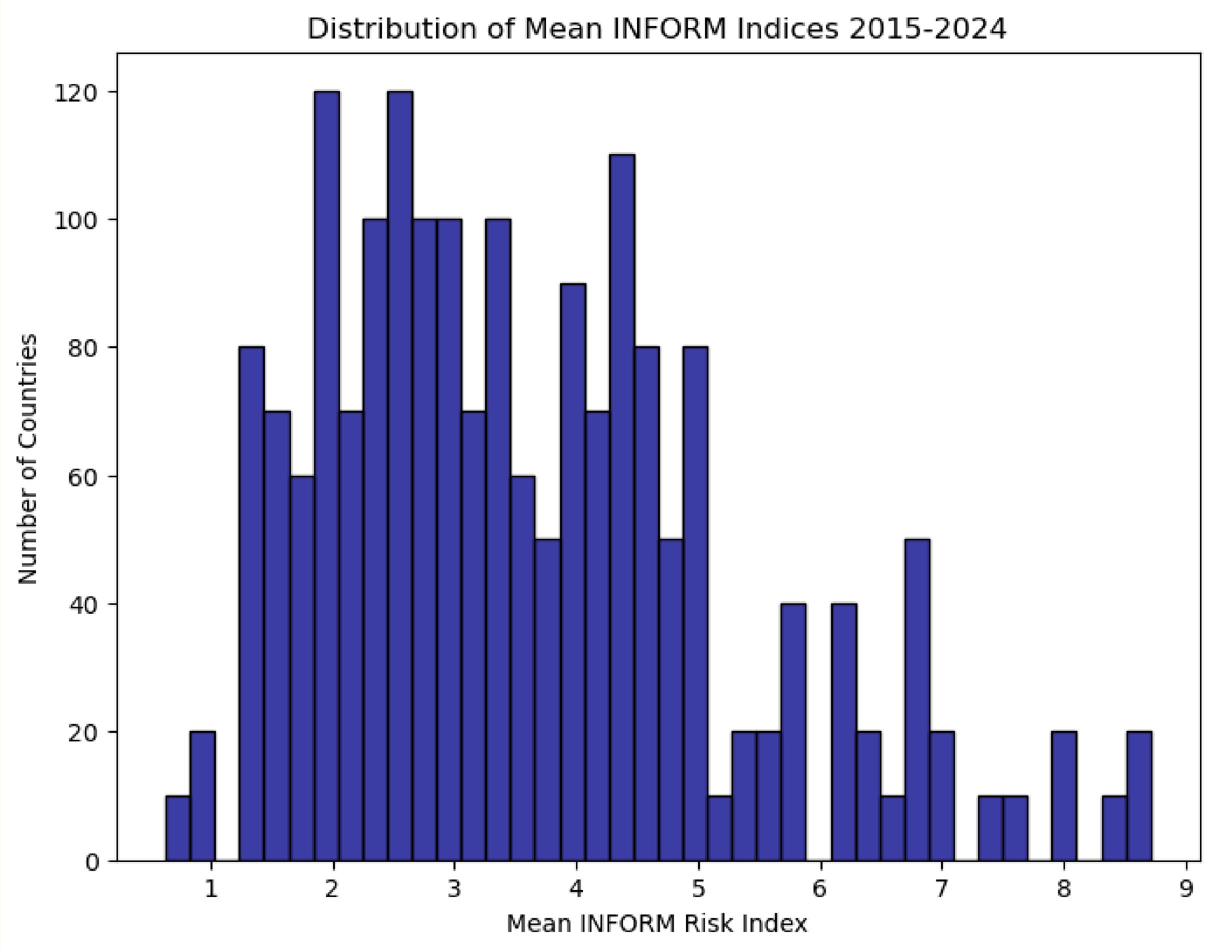
Duplicate Year column before setting as dataframe Index for easy groupby

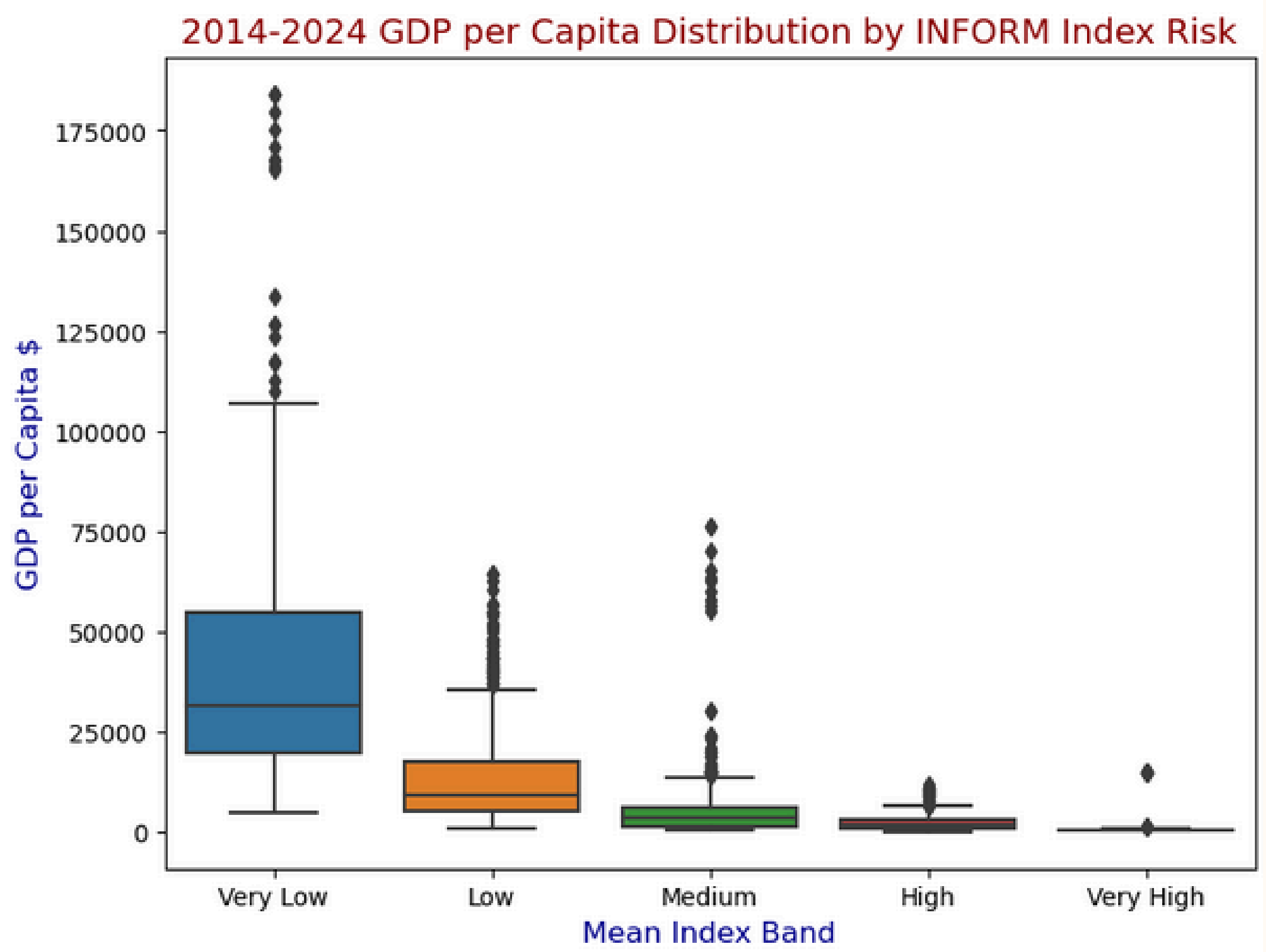
EDA: 2015-2024 INFORM Risk Index Categories



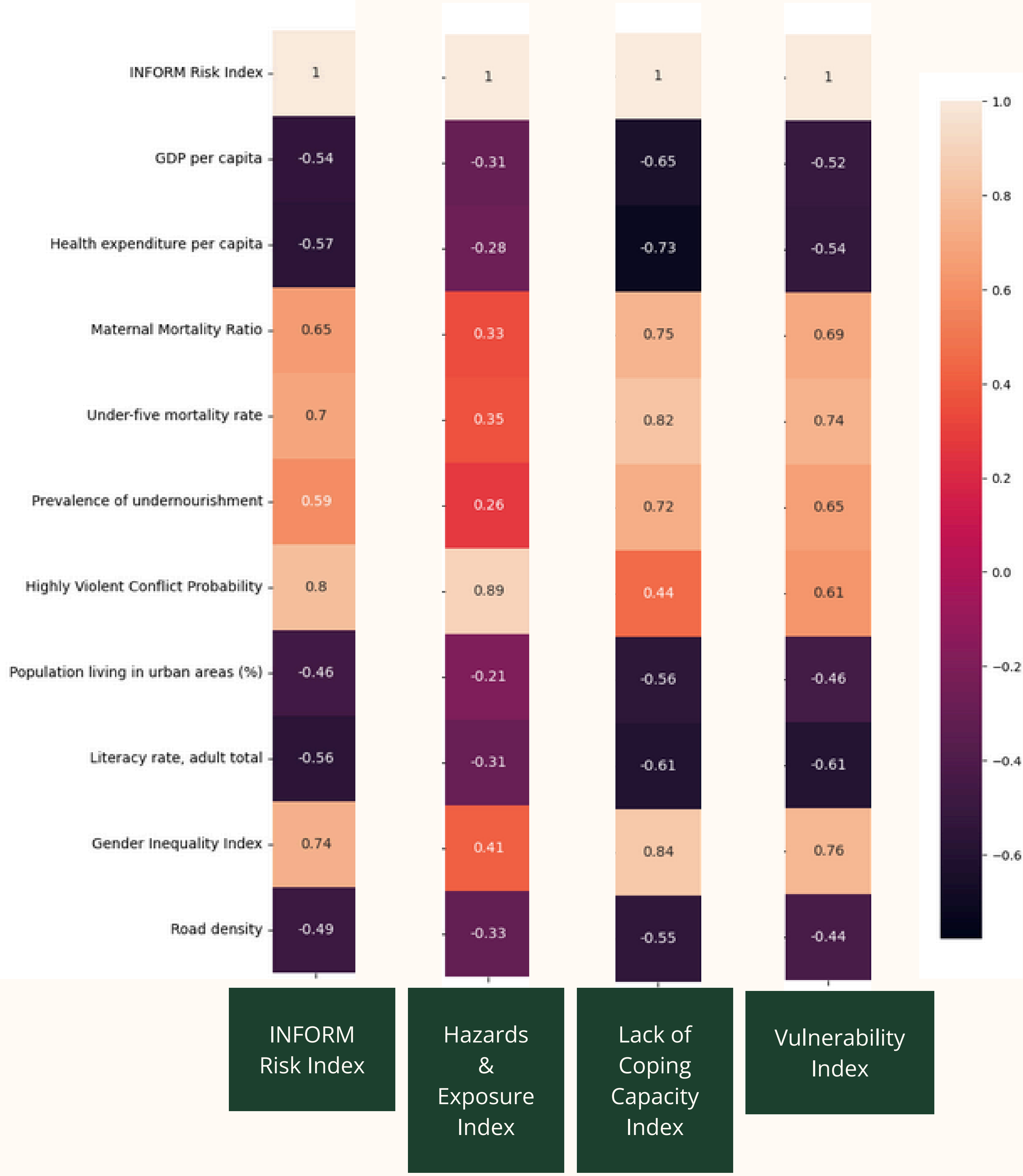
EDA

Back to Agenda



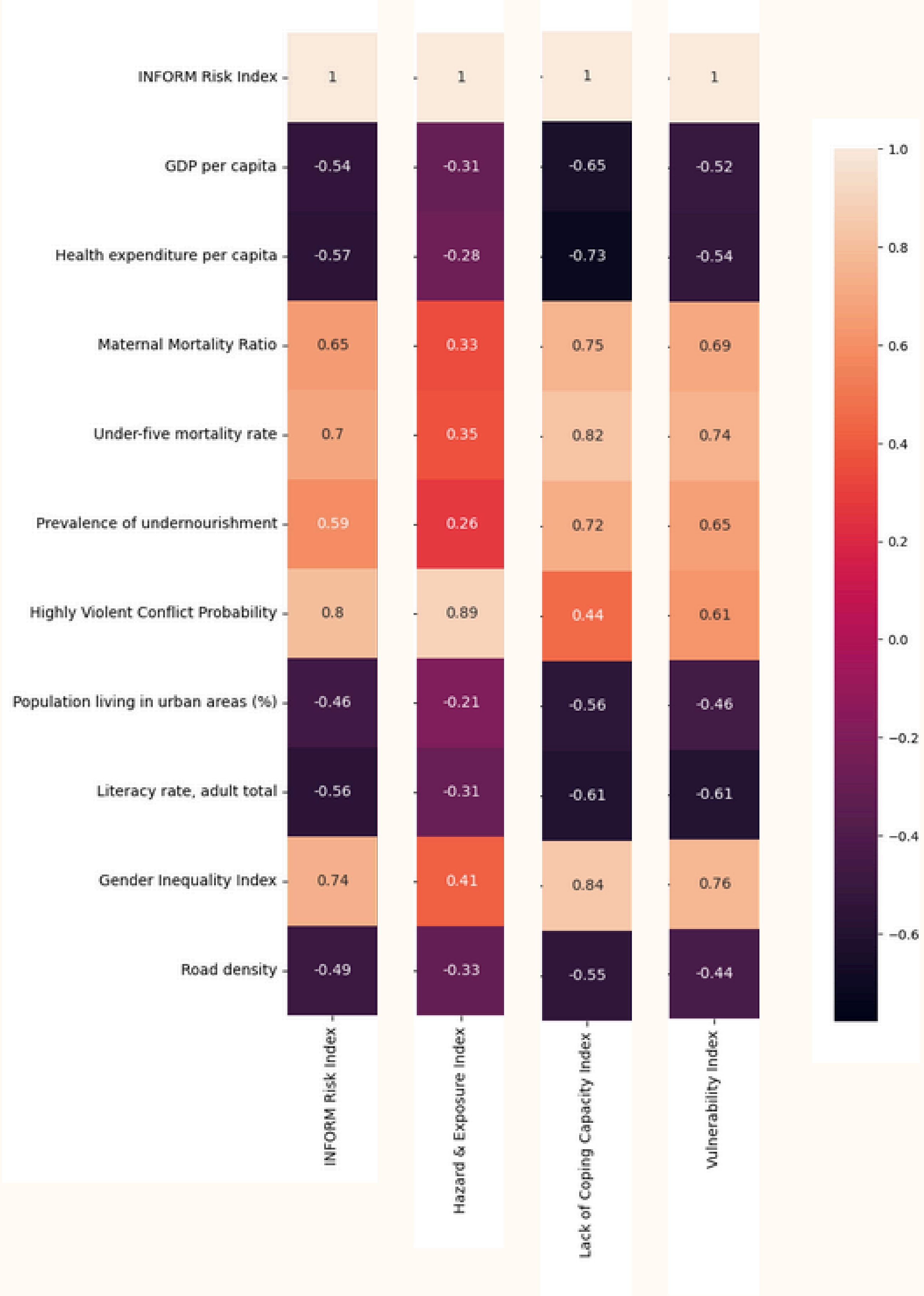


INFORM Indices vs Core Indicators

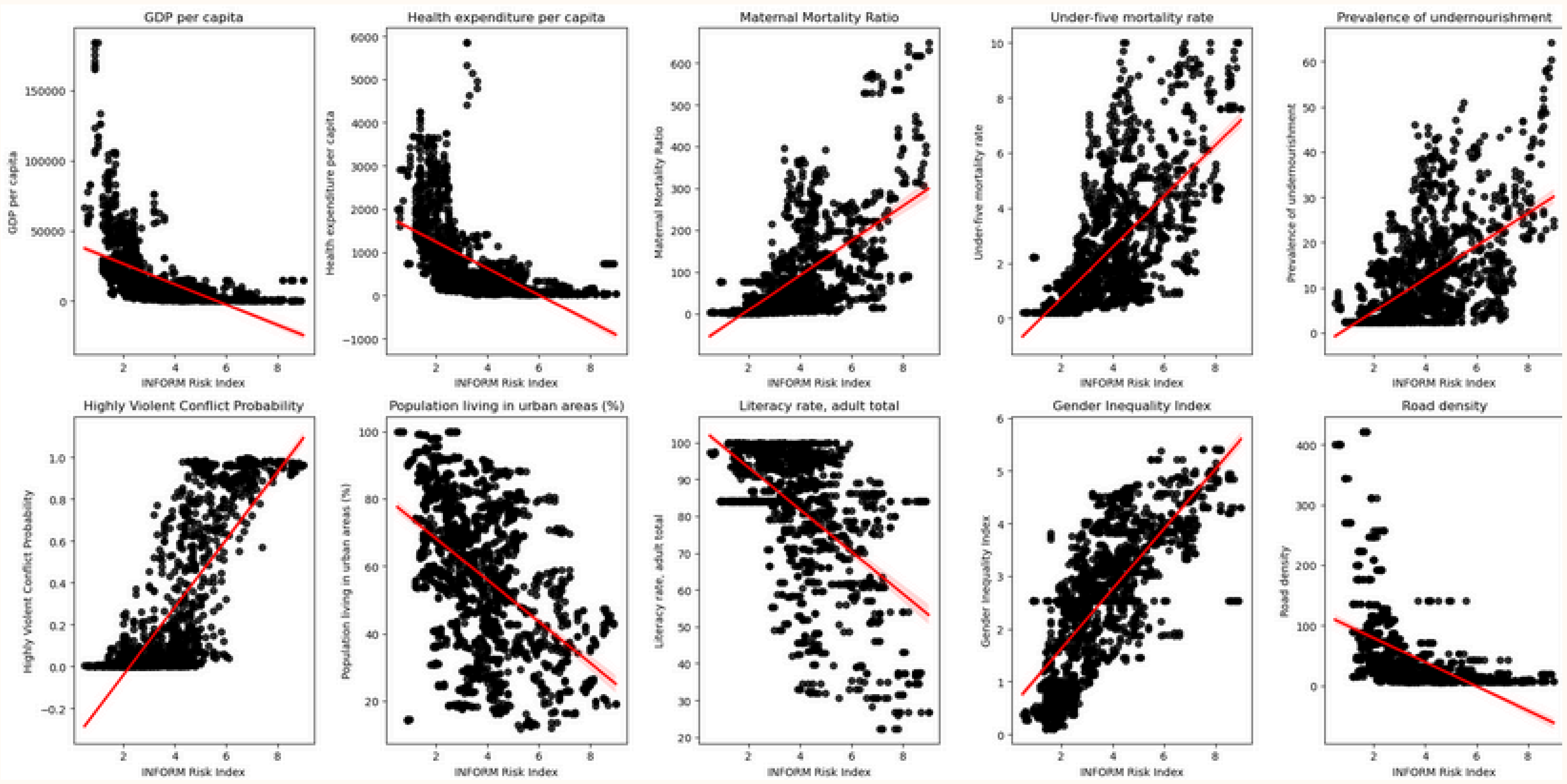


EDA

Back to Agenda



Correlations: INFORM Index vs Selected Core Indicators



Back to Agenda

Project 3

Predict & forecast how INFORM risk data changes over time

Machine Learning

ML Models

Can we forecast the INFORM Risk index?

Time Series Models for Prediction & Forecasting

1. ARIMA

Problem #3

Back to Agenda

Can we forecast the INFORM Risk index?

- Machine Learning: Time Series Models
  - Traditional Models
    - **Univariate**
      - ARIMA / SARIMA
      - Prophet / Neural Prohet
    - **Multivariate**
      - Vector autoregression
  - Other Models:
    - Long Short-Term Memory (LSTM)
    - Other regressor

Problem #3

Back to Agenda

AutoRegressive (AR)
Integrated (I)
Moving Average (MA)
(ARIMA)

AR(p) AutoRegression

- Regression model that utilizes the dependent relationship between a current observation & observations over a previous period
- An auto regressive (AR(p)) component refers to the use of past values in the regression equation for the time series)

I(d) Integration

- Uses differencing of observations (subtracting an observation from observation at the previous time step) in order to make the time series stationary.
- Differencing involves the subtraction of the current values of a series with its previous values d number of times.

MA(q) Moving Average

- A model that uses the dependency between an observation & a residual error from a moving average model applied to lagged observations. A moving average component depicts the error of the model as a combination of previous error terms. The order q represents the number of terms to be included in the model.

Problem #3

Back to Agenda

AutoRegressive (AR)
Integrated (I)
Moving Average (MA)
(ARIMA)

Checking Autocorrelation to determine AR(p)

- Partial Autocorrelation conveys pure correlation of a series & its lag, excluding correlation contributions from intermediate lags.

Test for Stationarity - I(d) Integration

- Uses differencing of observations (subtracting an observation from observation at the previous time step) to make time series stationary
- Subtraction of current values of a series with its previous values d number of times
- Stationary time series has statistical properties or moments (e.g., mean & variance) that do not vary in time. Stationarity is status of a stationary time series. Conversely, nonstationarity - the status of a time series whose statistical properties are changing through time

Checking Autocorrelation to determine MA(q)

- Autocorrelation is the correlation of a series with its own lags.
- If a series is significantly autocorrelated, that means previous values of series (lags) may be helpful in predicting the current value

Problem #3

Back to Agenda

AutoRegressive (AR)
Integrated (I)
Moving Average (MA)
(ARIMA)

1. **Autocorrelation Function (ACF) Plot:**
   - ACF plot helps identify "q" parameter (order of moving average component)
   - If ACF plot shows a significant spike at lag "q", then cuts off abruptly or decreases gradually, suggests presence of a moving average component of order "q".
2. **Partial Autocorrelation Function (PACF) Plot:**
   - PACF plot helps identify the "p" parameter (order of autoregressive component)
   - If PACF plot shows a significant spike at lag "p", then cuts off abruptly or decreases gradually, suggests presence of an autoregressive component of order "p".

Interpretation:

- If both ACF & PACF plots show significant spikes at multiple lags, might suggest more complex ARIMA model with both autoregressive & moving average components (i.e., an ARIMA(p, d, q) model).
- If ACF plot decays exponentially & PACF plot cuts off after a few lags, it may indicate that series can be modeled with a pure moving average process of order "q" (i.e., an MA(q) model).
- Conversely, if PACF plot decays exponentially & ACF plot cuts off after a few lags, may suggest series can be modeled with a pure autoregressive process of order "p" (i.e., an AR(p) model).

Problem #3

Back to Agenda

ARIMA Output

- **Mean Squared Error (MSE):** MSE measures average squared difference between the actual & predicted values. It gives more weight to large errors because of the squaring operation. Lower MSE values indicate better model performance.

- **Mean Absolute Error (MAE):** MAE measures average absolute difference between the actual & predicted values. It treats all errors equally regardless of their direction (positive or negative). Where the symbols have the same meaning as in MSE. Like MSE, lower MAE values indicate better model performance.

- **Root Mean Squared Error (RMSE):** RMSE is simply the square root of MSE. It provides an interpretable measure of the average magnitude of the error, in the same units as the response variable. Since RMSE is in the same units as the response variable, it's easier to interpret compared to MSE.

- **ar.L1 & ma.L1** -> estimated coefficients of autoregressive (AR) & moving average (MA)
- **sigma2** -> estimated variance of error term

Problem #3

ARIMA Output

Log Likelihood: Goodness of fit of model to data. Higher log likelihood values indicate better fit of model to data.
AIC (Akaike Information Criterion: Measures relative model quality for data. Penalizes model for complexity, helps select model that most effectively explains data while using fewest parameters. Lower AIC values indicate a better trade-off between goodness of fit & model complexity.

BIC (Bayesian Information Criterion): Penalizes the model for complexity. Puts higher penalty on models with more parameters than AIC. Lower BIC values -> better trade-off between goodness of fit & model complexity.

HQIC (Hannan-Quinn Information Criterion): Another criterion for model selection, similar to AIC & BIC, penalizes model for complexity. Lower HQIC values indicate better trade-off between goodness of fit & model complexity

- Lower AIC, BIC, or HQIC value suggests better-fitting model
- Higher log likelihood values -> better fit
- Compare these values across models to determine better fit model

- **Ljung-Box (L1) (Q):** tests whether any autocorrelation in residuals is different from zero at lag 1. Low p-value suggests evidence of autocorrelation in residuals at lag 1.
- **Jarque-Bera (JB):** tests null hypothesis that data is normally distributed. Low p-value -> residuals not normally distributed
- **Heteroskedasticity (H):** Tests null hypothesis of homoscedasticity (constant variance of errors). Low p-value suggests evidence against null hypothesis -> heteroskedasticity in residuals
- **Skew & Kurtosis:** Measures distribution - skewness & kurtosis/ "tailedness" of residuals

Back to Agenda

Project 3

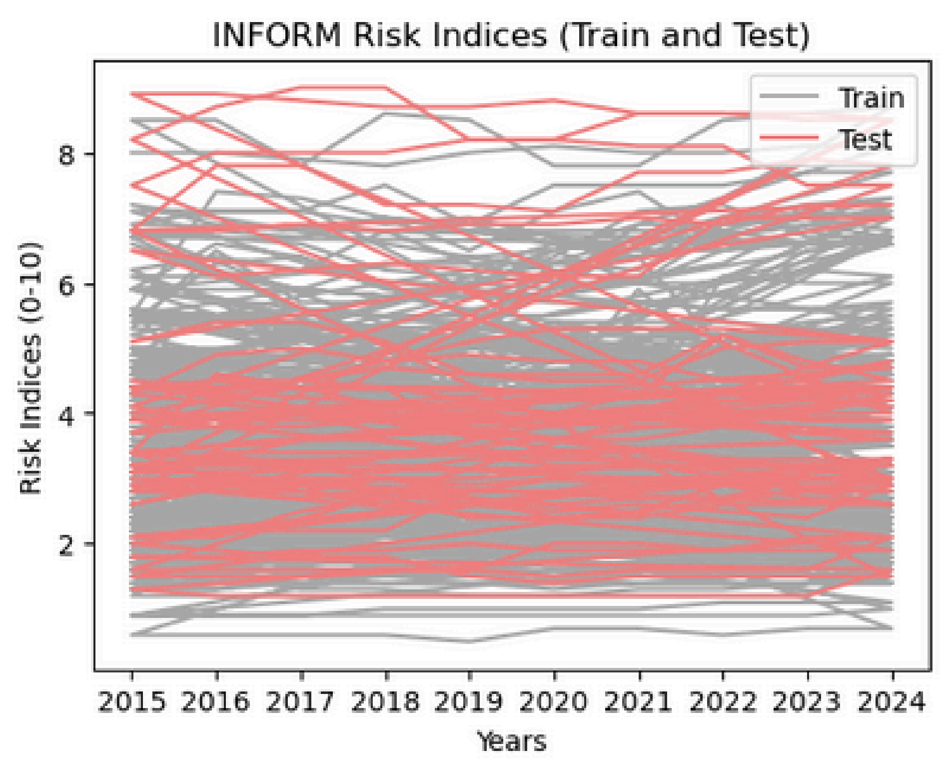
Predict & forecast how INFORM risk data changes over time

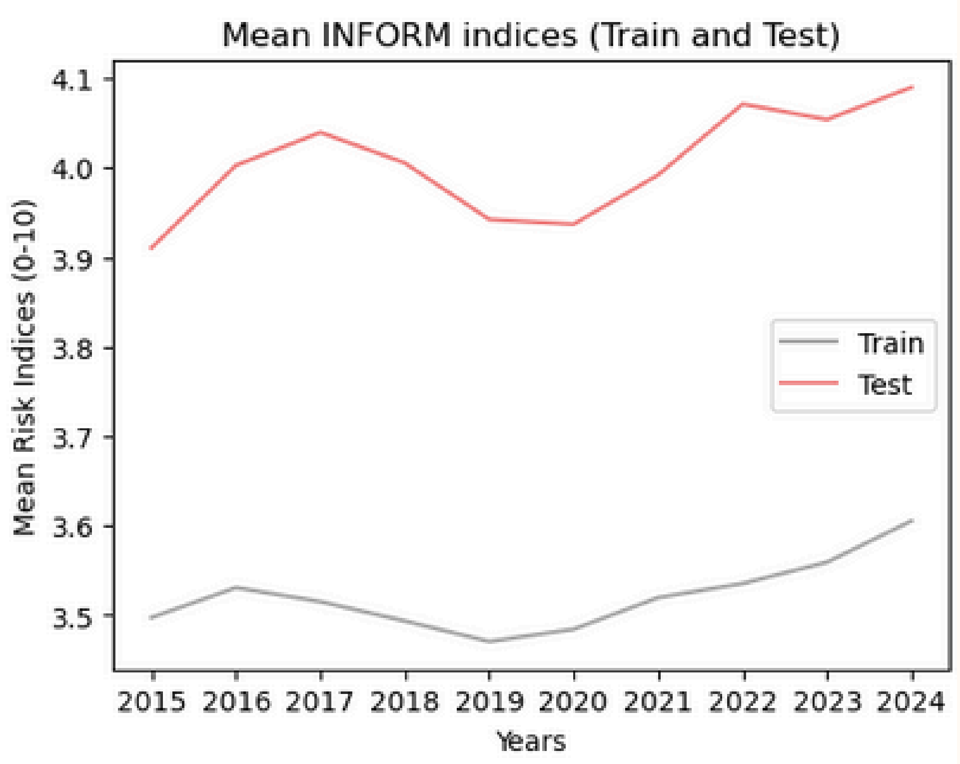
INFORM Risk Index

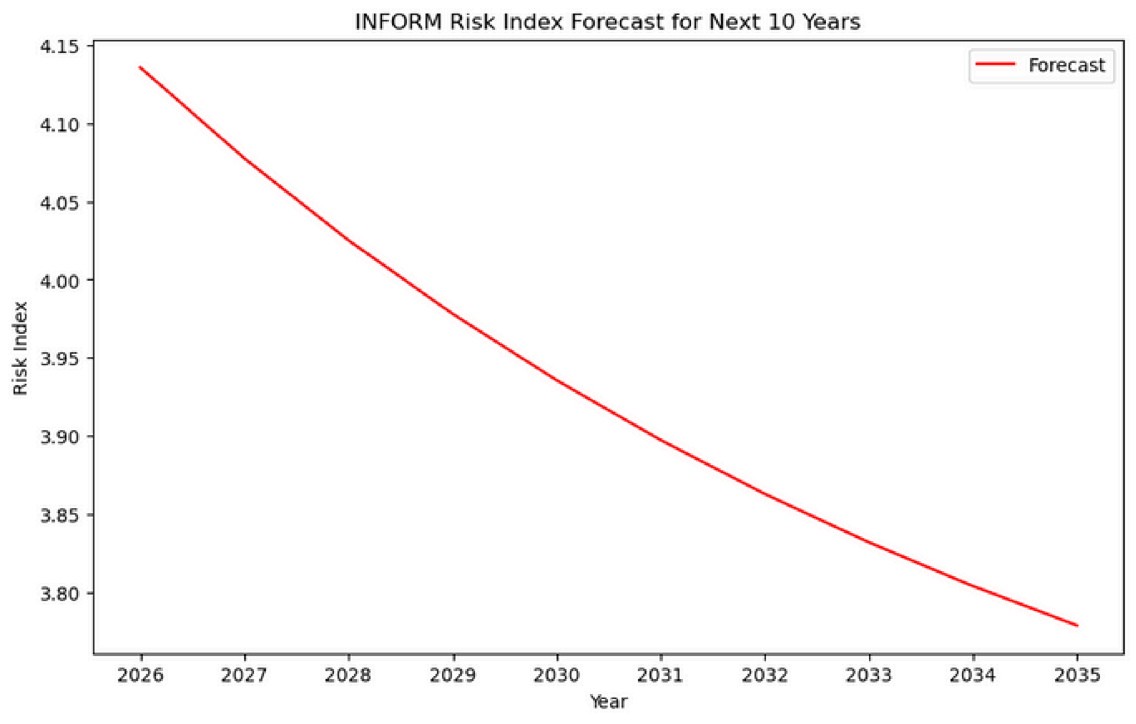
ARIMA Model

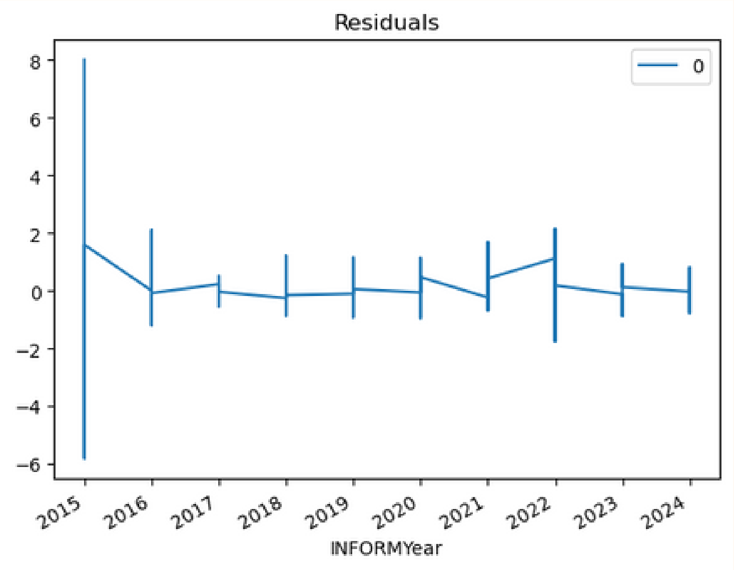
ARIMA Train - Test Model to forecast INFORM Risk indices:

Back to Agenda







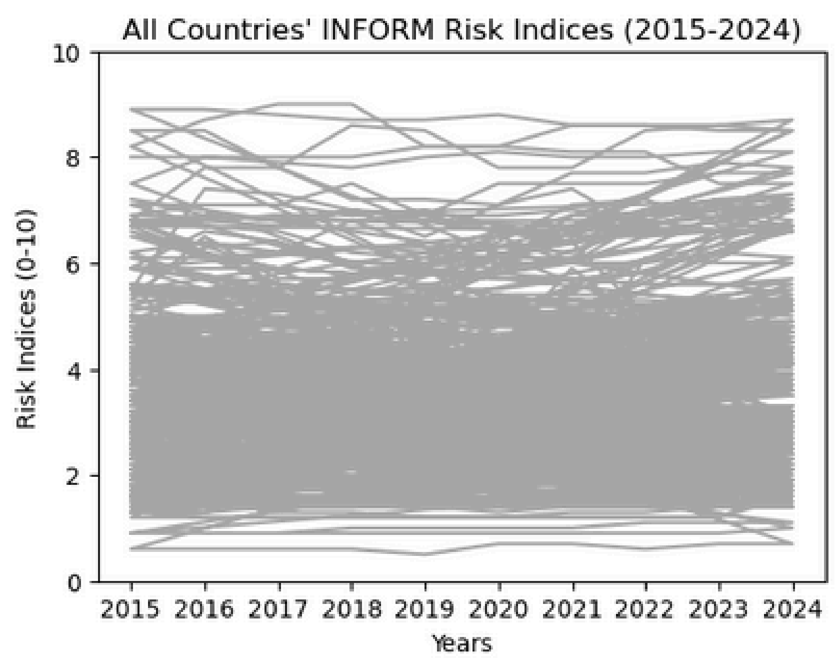


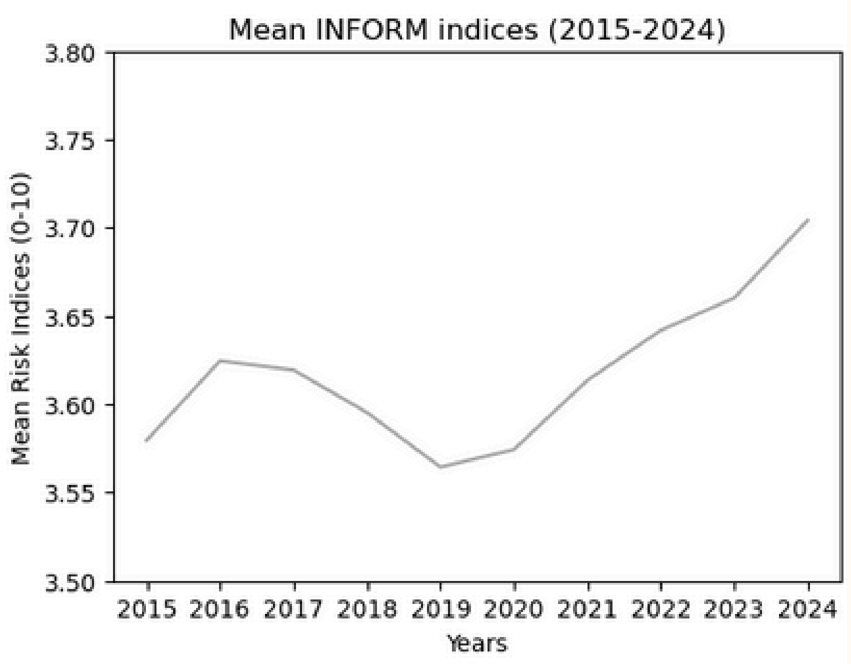

```
Mean Squared Error (MSE): 3.6829821042967064
Mean Absolute Error (MAE): 1.377148768592445
Root Mean Squared Error (RMSE): 1.9191097165864974
```

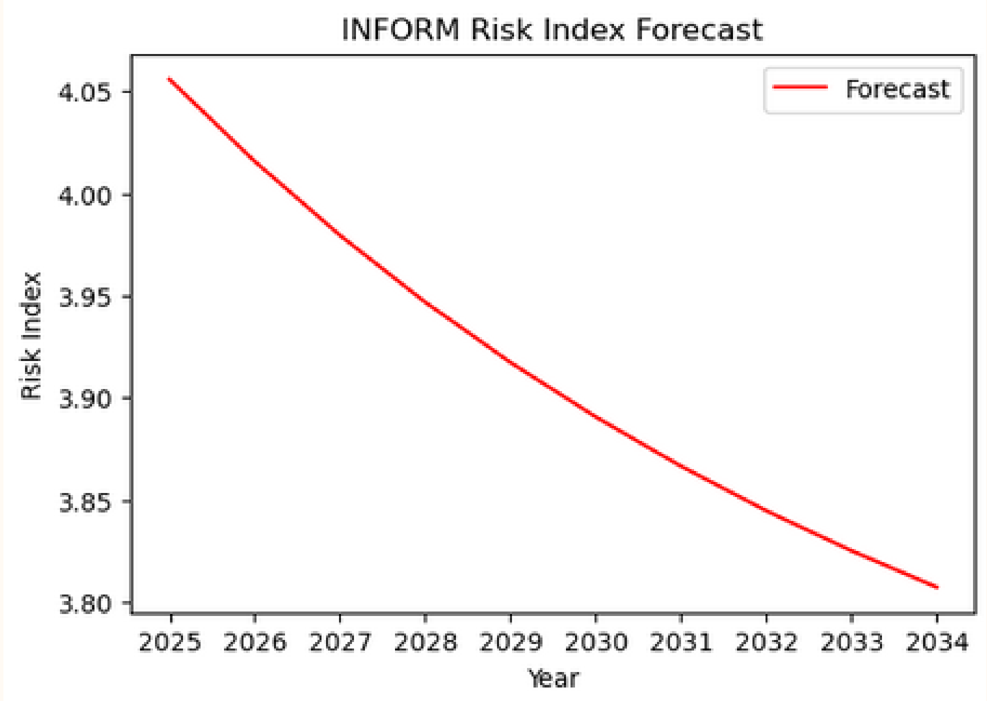
```
                               SARIMAX Results
==============================================================================
Dep. Variable:      INFORM Risk Index   No. Observations:                 1528
Model:                 ARIMA(1, 1, 1)   Log Likelihood               -1671.588
Date:                Mon, 08 Apr 2024   AIC                           3349.176
Time:                        21:39:47   BIC                           3365.169
Sample:                             0   HQIC                          3355.129
                               - 1528
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.9007      0.013     66.950      0.000       0.874       0.927
ma.L1         -1.0000      0.245     -4.077      0.000      -1.481      -0.519
sigma2         0.5213      0.129      4.034      0.000       0.268       0.775
===================================================================================
Ljung-Box (L1) (Q):                   3.09   Jarque-Bera (JB):             30368.06
Prob(Q):                              0.08   Prob(JB):                         0.00
Heteroskedasticity (H):               1.52   Skew:                            -0.73
Prob(H) (two-sided):                  0.00   Kurtosis:                        24.80
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

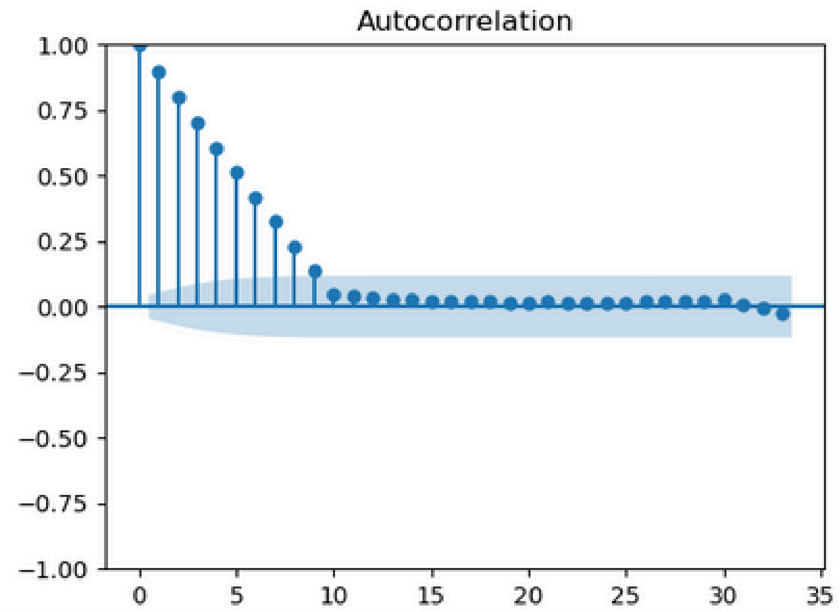
ARIMA Model to forecast INFORM Risk indices:

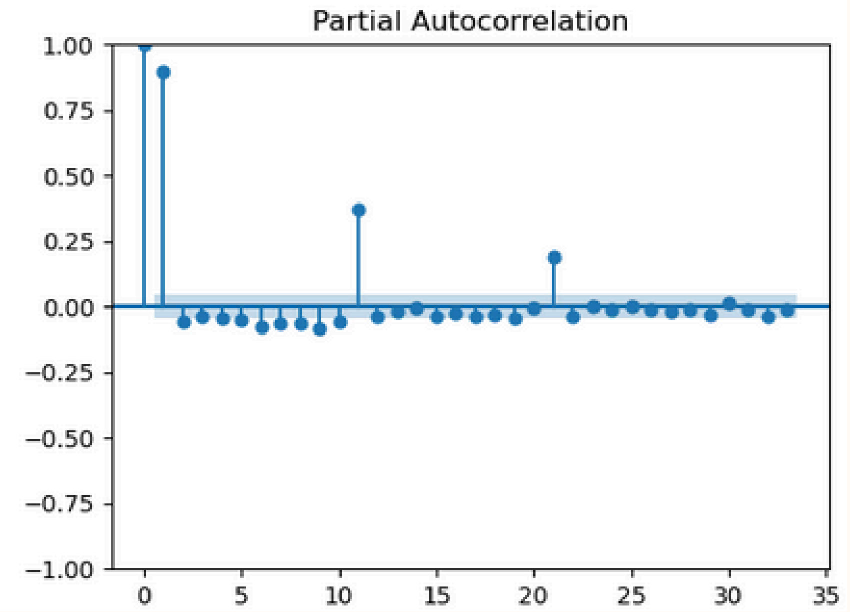
Back to Agenda

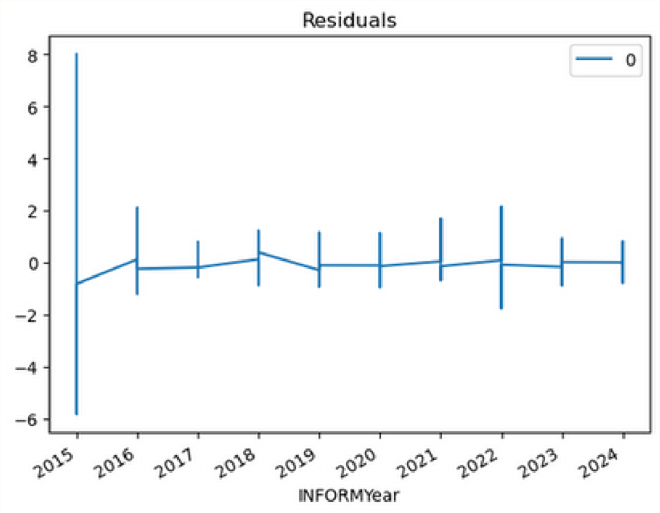










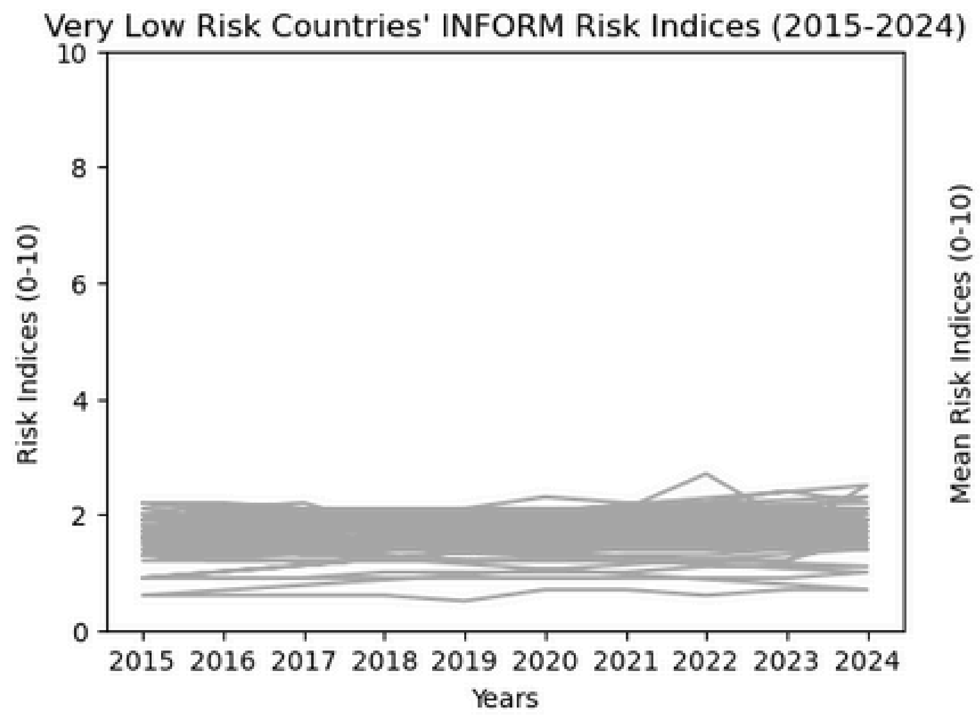


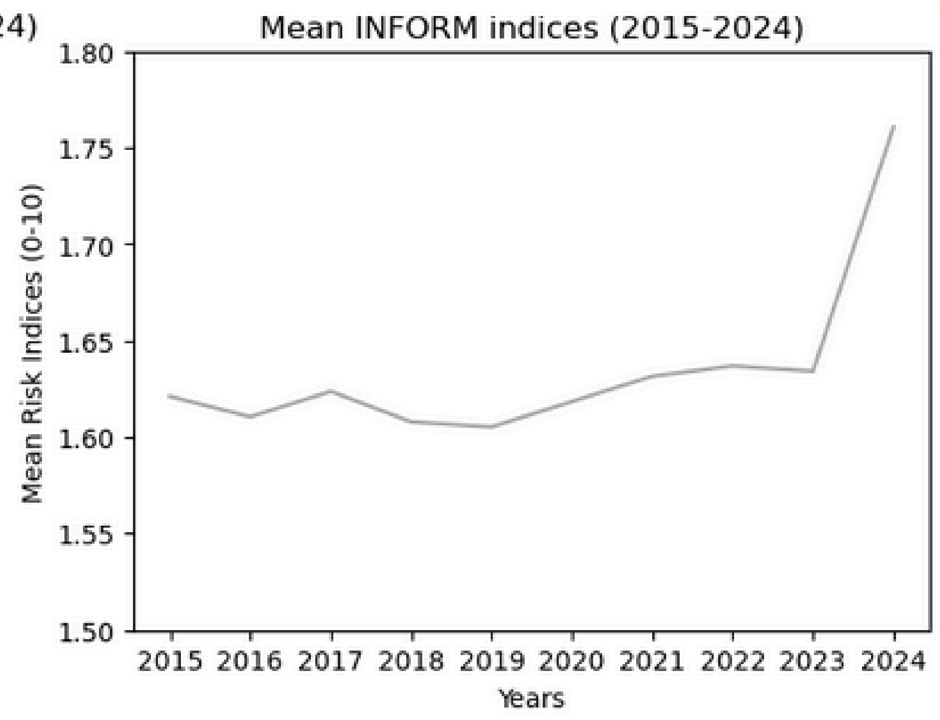

```
Mean Squared Error (MSE): 0.10595662427130725
Mean Absolute Error (MAE): 0.29619233818658486
Root Mean Squared Error (RMSE): 0.325509791360117
```

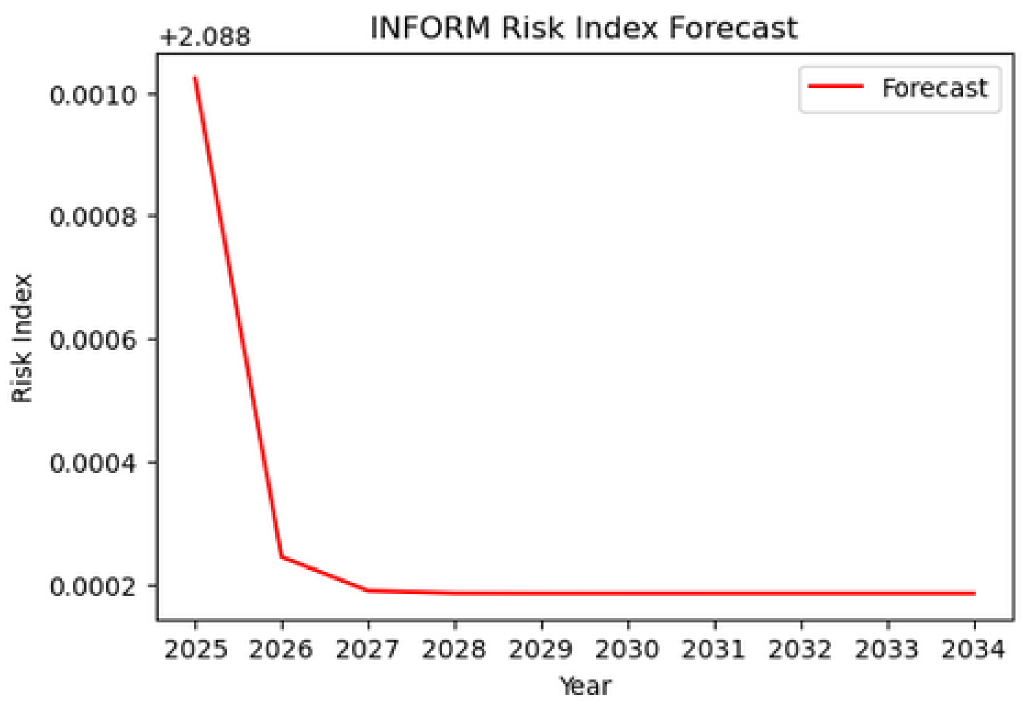
```
                               SARIMAX Results
==============================================================================
Dep. Variable:     INFORM Risk Index   No. Observations:                 1910
Model:                ARIMA(1, 1, 1)   Log Likelihood               -2128.094
Date:               Mon, 08 Apr 2024   AIC                           4262.187
Time:                       02:50:43   BIC                           4278.850
Sample:                            0   HQIC                          4268.320
                              - 1910
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.9033      0.012     73.023      0.000       0.879       0.928
ma.L1         -0.9999      0.059    -16.983      0.000      -1.115      -0.885
sigma2         0.5430      0.030     17.965      0.000       0.484       0.602
===================================================================================
Ljung-Box (L1) (Q):                   4.56   Jarque-Bera (JB):             41753.34
Prob(Q):                              0.03   Prob(JB):                         0.00
Heteroskedasticity (H):               1.35   Skew:                            -0.44
Prob(H) (two-sided):                  0.00   Kurtosis:                        25.89
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

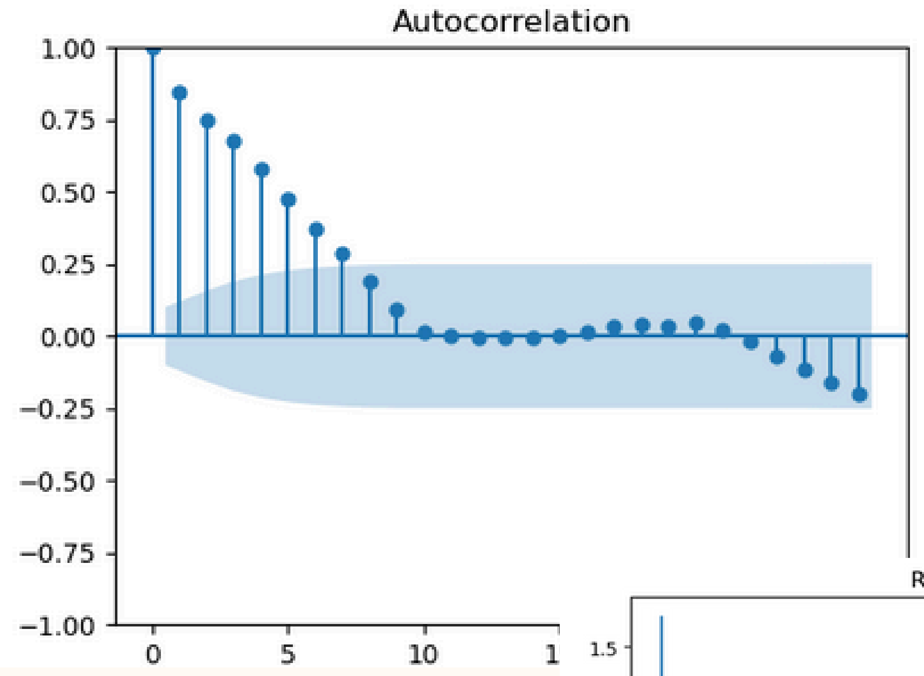
ARIMA Model to forecast Very Low Risk Countries' INFORM Risk indices:

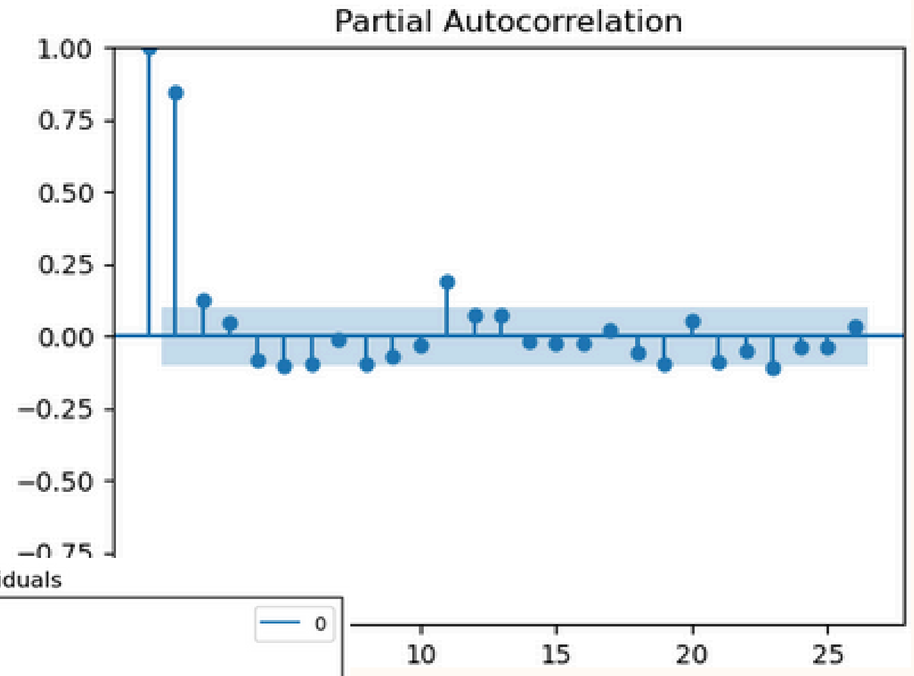
Back to Agenda

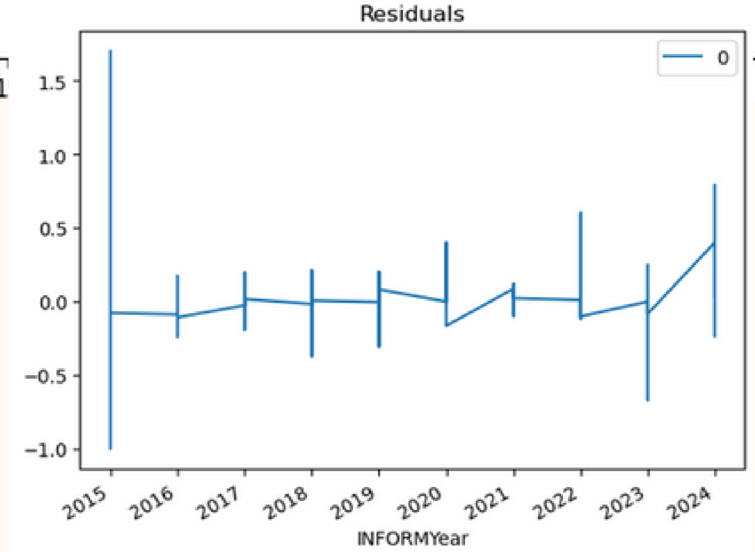










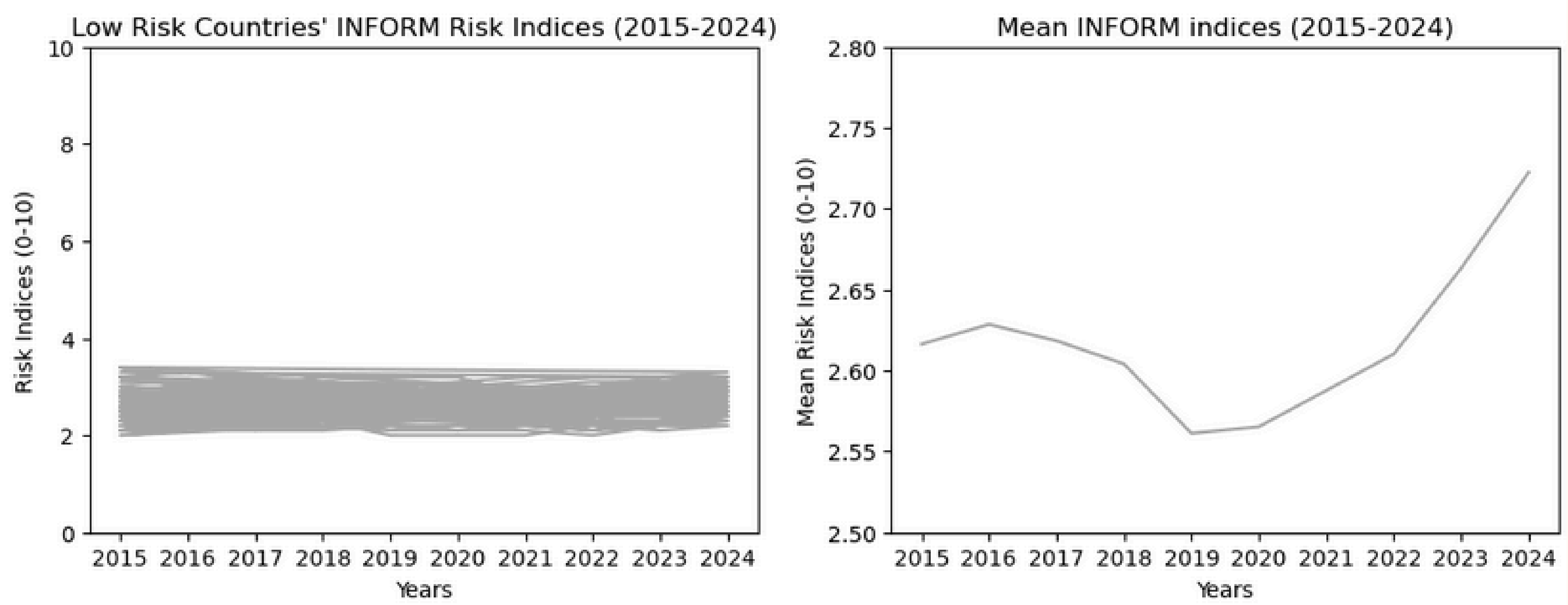


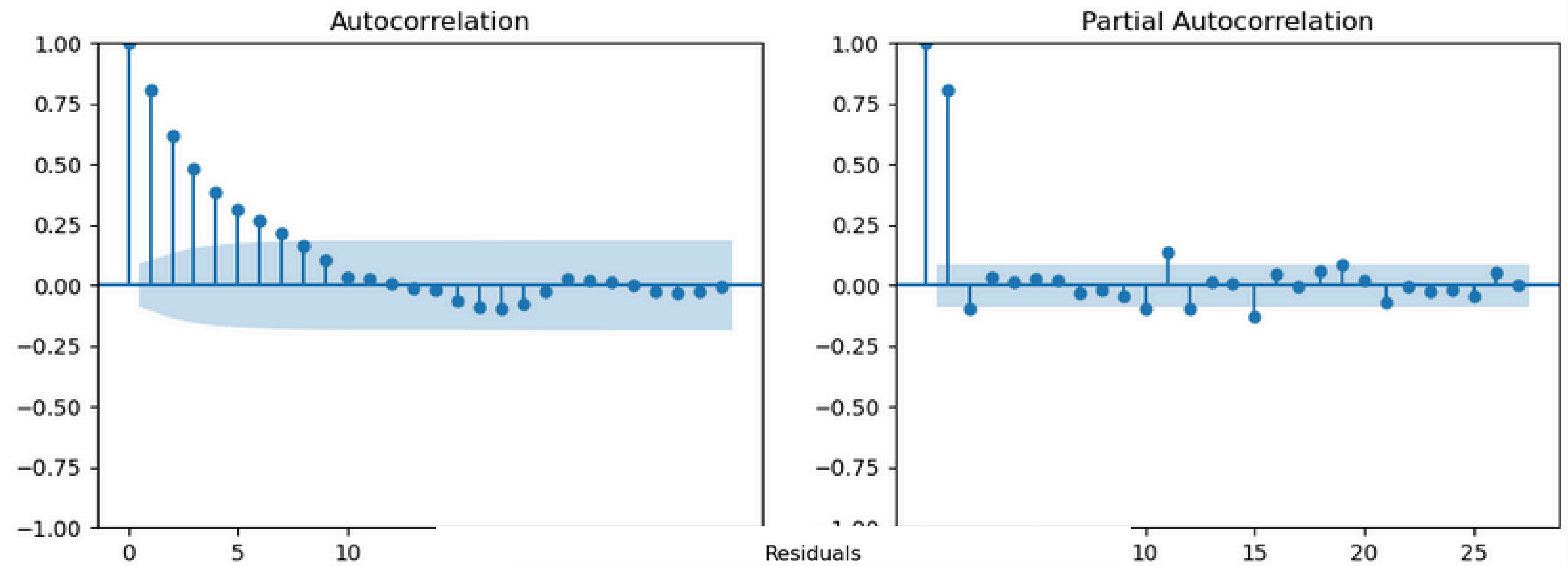

```
Mean Squared Error (MSE): 0.12437508160867335
Mean Absolute Error (MAE): 0.2830012484014993
Root Mean Squared Error (RMSE): 0.35266851519333753
```

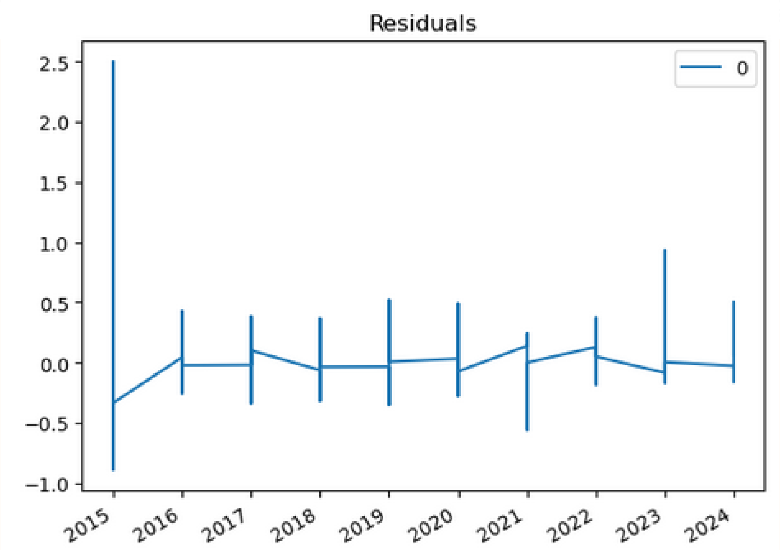
```
                               SARIMAX Results
==============================================================================
Dep. Variable:      INFORM Risk Index   No. Observations:                  380
Model:                 ARIMA(1, 1, 1)   Log Likelihood                  84.230
Date:                Mon, 08 Apr 2024   AIC                           -162.460
Time:                        03:19:44   BIC                           -150.648
Sample:                             0   HQIC                          -157.772
                                - 380
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0710      0.420      0.169      0.866      -0.752       0.894
ma.L1         -0.2844      0.424     -0.670      0.503      -1.116       0.547
sigma2         0.0375      0.001     27.819      0.000       0.035       0.040
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):               997.86
Prob(Q):                              0.98   Prob(JB):                         0.00
Heteroskedasticity (H):               1.58   Skew:                            -0.49
Prob(H) (two-sided):                  0.01   Kurtosis:                        10.89
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

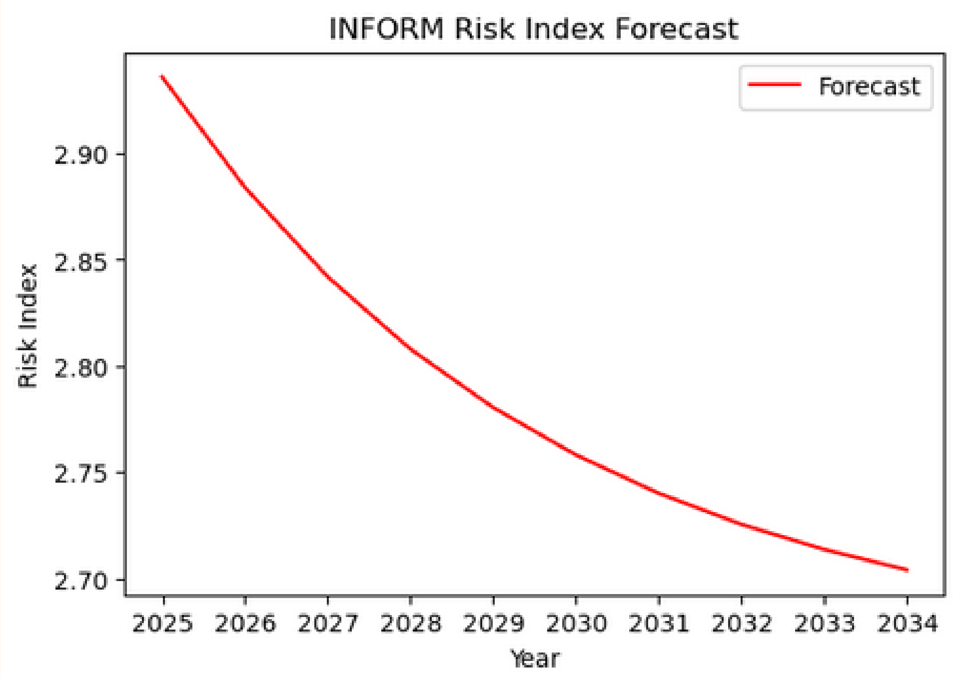
ARIMA Model to forecast Low Risk Countries' INFORM Risk indices:

Back to Agenda







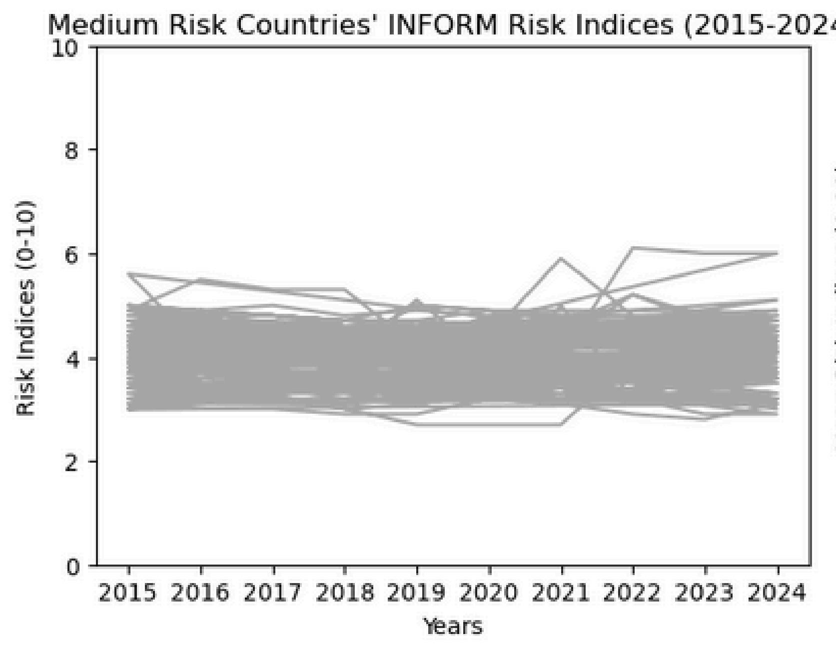


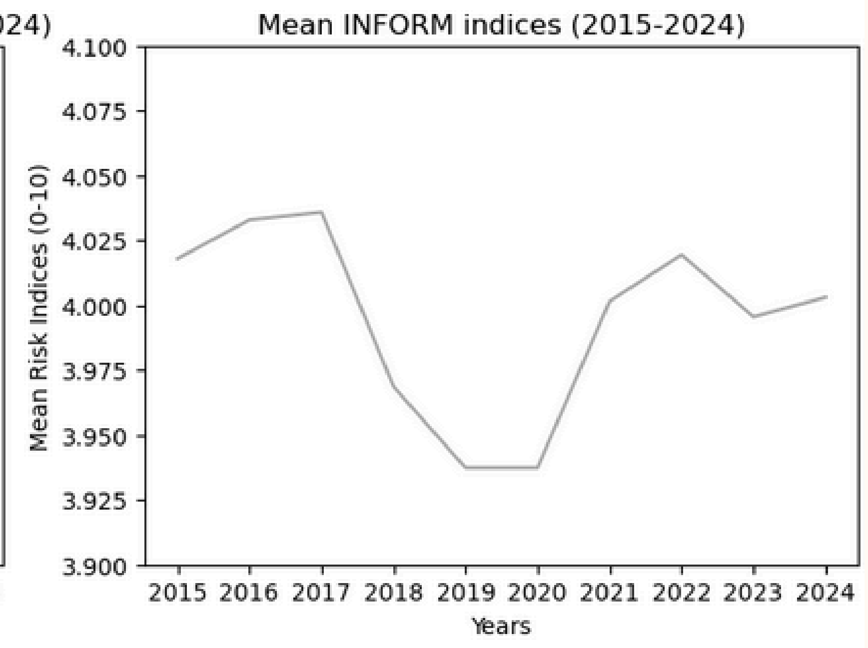

```
Mean Squared Error (MSE): 0.040013193774869 52
Mean Absolute Error (MAE): 0.16483573517531735
Root Mean Squared Error (RMSE): 0.20003298171768955
```

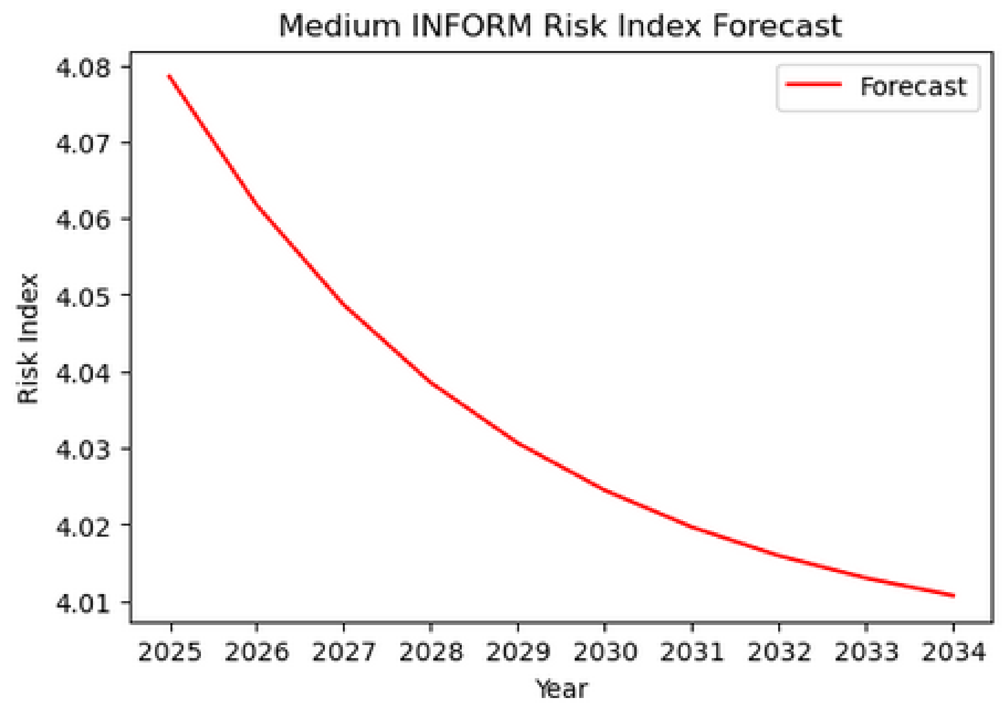
```
                               SARIMAX Results
==============================================================================
Dep. Variable:      INFORM Risk Index   No. Observations:                  490
Model:                 ARIMA(1, 1, 1)   Log Likelihood                 131.461
Date:                Mon, 08 Apr 2024   AIC                           -256.922
Time:                        03:26:37   BIC                           -244.344
Sample:                             0   HQIC                          -251.982
                                - 490
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8097      0.033     24.459      0.000       0.745       0.875
ma.L1         -0.9952      0.011    -94.641      0.000      -1.016      -0.975
sigma2         0.0340      0.001     22.974      0.000       0.031       0.037
===================================================================================
Ljung-Box (L1) (Q):                   2.81   Jarque-Bera (JB):               727.19
Prob(Q):                              0.09   Prob(JB):                         0.00
Heteroskedasticity (H):               0.90   Skew:                            -0.37
Prob(H) (two-sided):                  0.49   Kurtosis:                         8.93
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

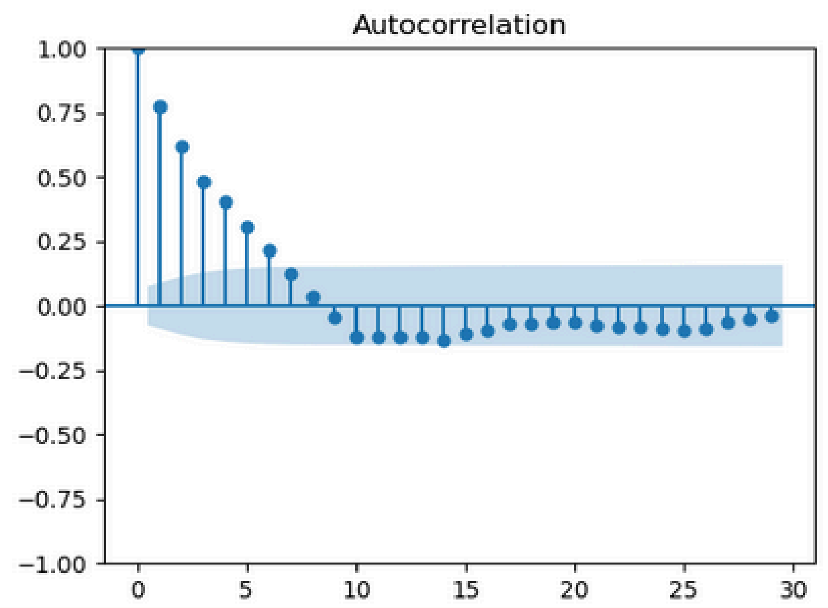
ARIMA Model to forecast Medium Risk Countries' INFORM Risk indices:

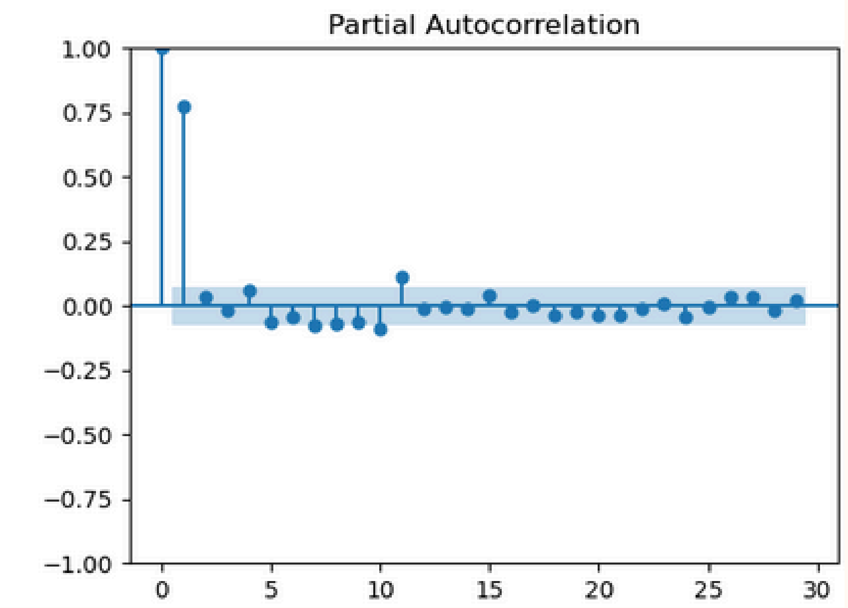
Back to Agenda

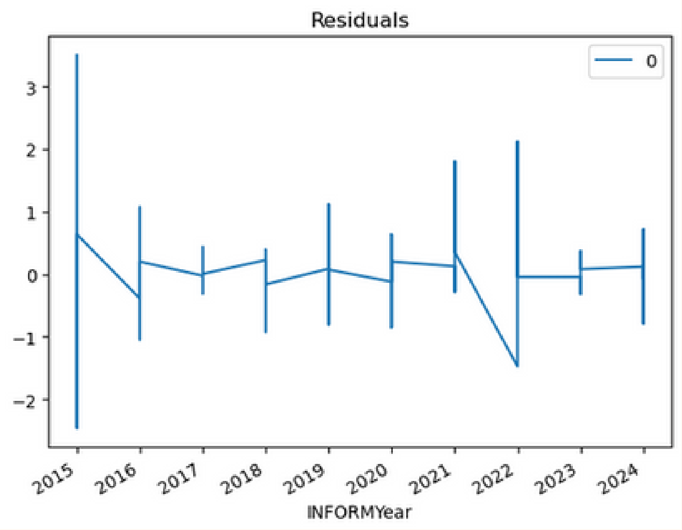










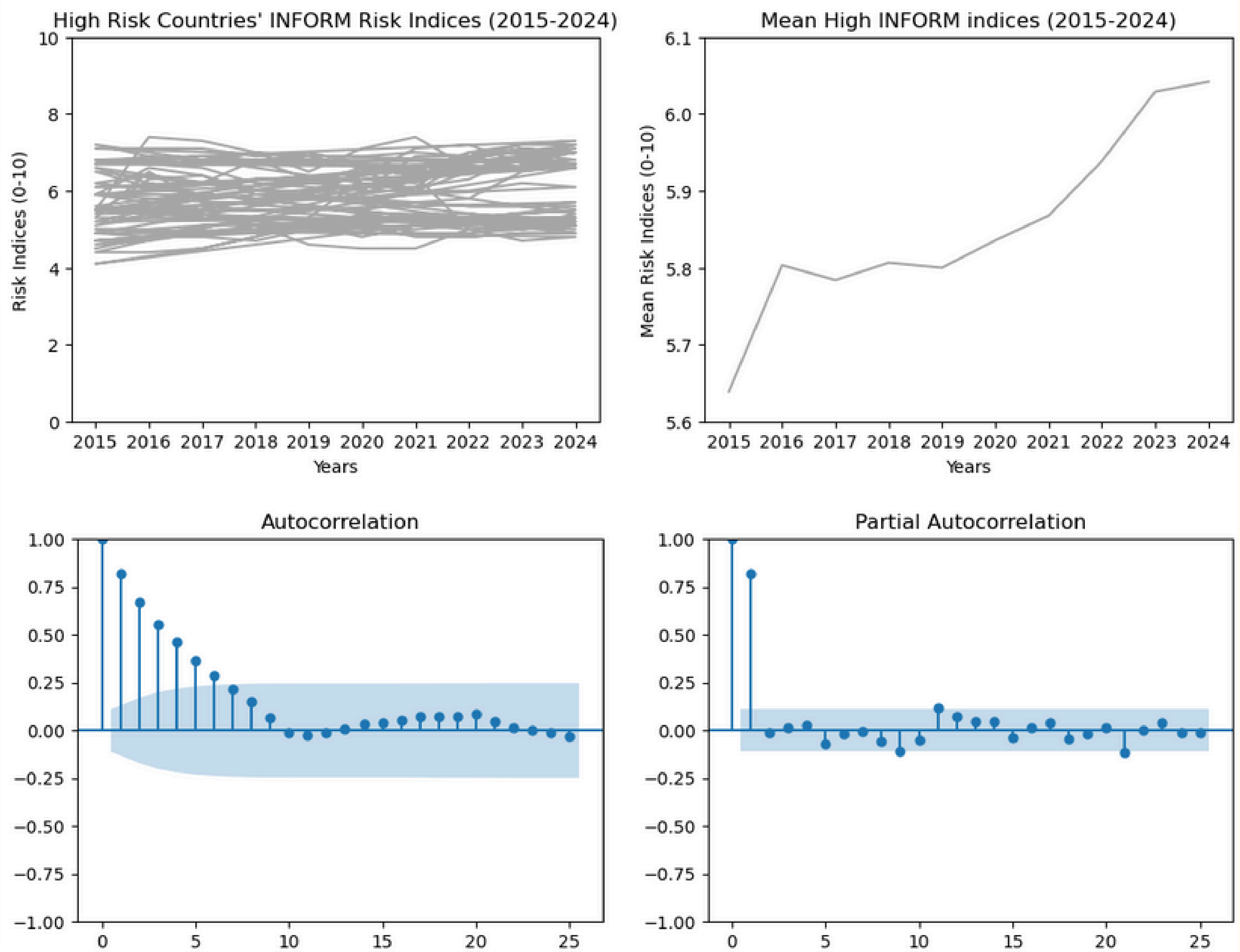


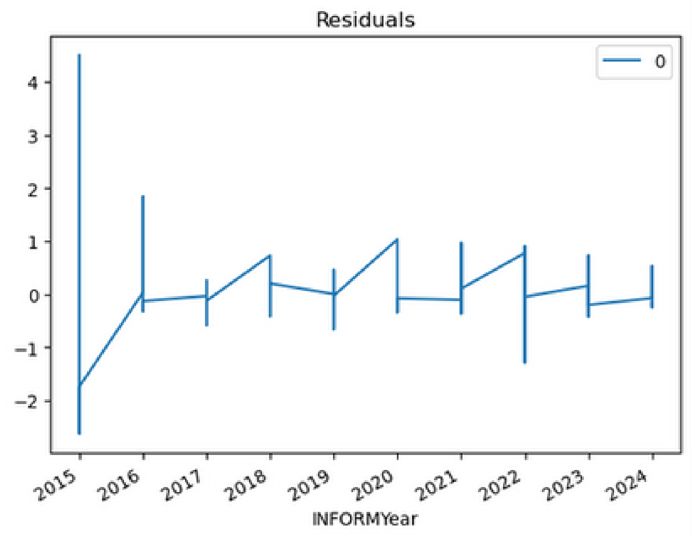

```
Mean Squared Error (MSE): 0.04353991388008316
Mean Absolute Error (MAE): 0.18150402194164225
Root Mean Squared Error (RMSE): 0.2086622004101565
```

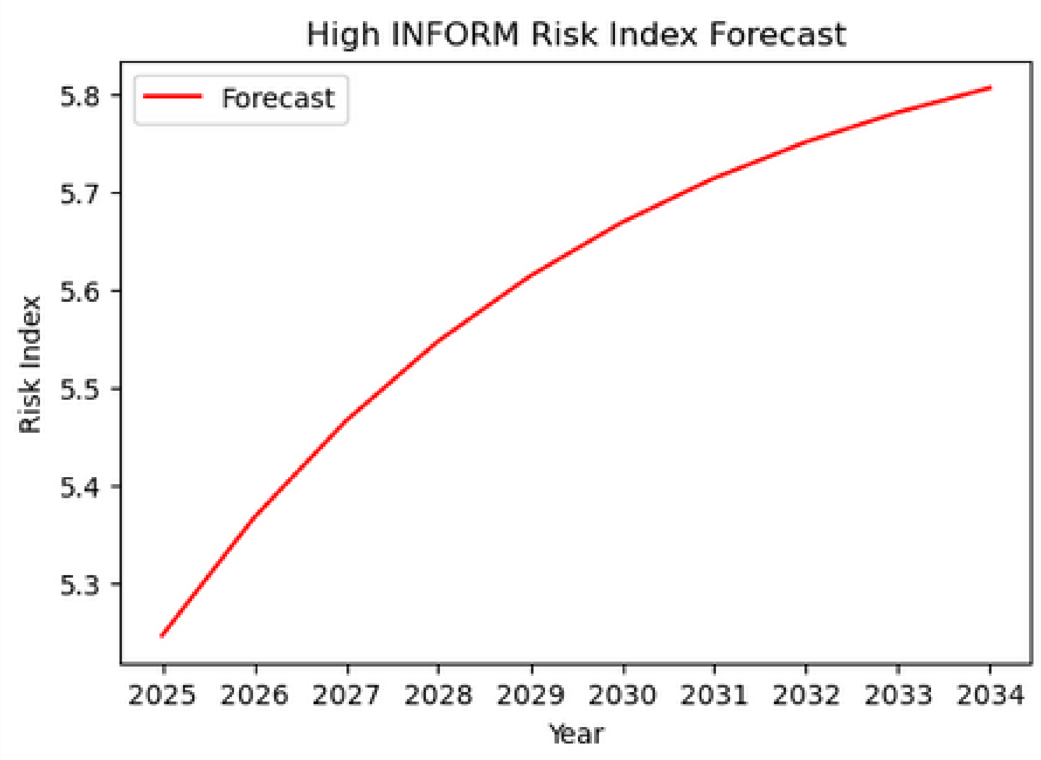
```
                               SARIMAX Results
==============================================================================
Dep. Variable:      INFORM Risk Index   No. Observations:                  670
Model:                 ARIMA(1, 1, 1)   Log Likelihood                -256.884
Date:                Mon, 08 Apr 2024   AIC                            519.769
Time:                        03:31:16   BIC                            533.286
Sample:                             0   HQIC                           525.005
                                - 670
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.7791      0.020     38.624      0.000       0.740       0.819
ma.L1         -0.9983      0.011    -94.869      0.000      -1.019      -0.978
sigma2         0.1255      0.003     36.389      0.000       0.119       0.132
===================================================================================
Ljung-Box (L1) (Q):                   0.78   Jarque-Bera (JB):              3175.76
Prob(Q):                              0.38   Prob(JB):                         0.00
Heteroskedasticity (H):               0.64   Skew:                             0.49
Prob(H) (two-sided):                  0.00   Kurtosis:                        13.63
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

ARIMA Model to forecast High Risk Countries' INFORM Risk indices:

Back to Agenda





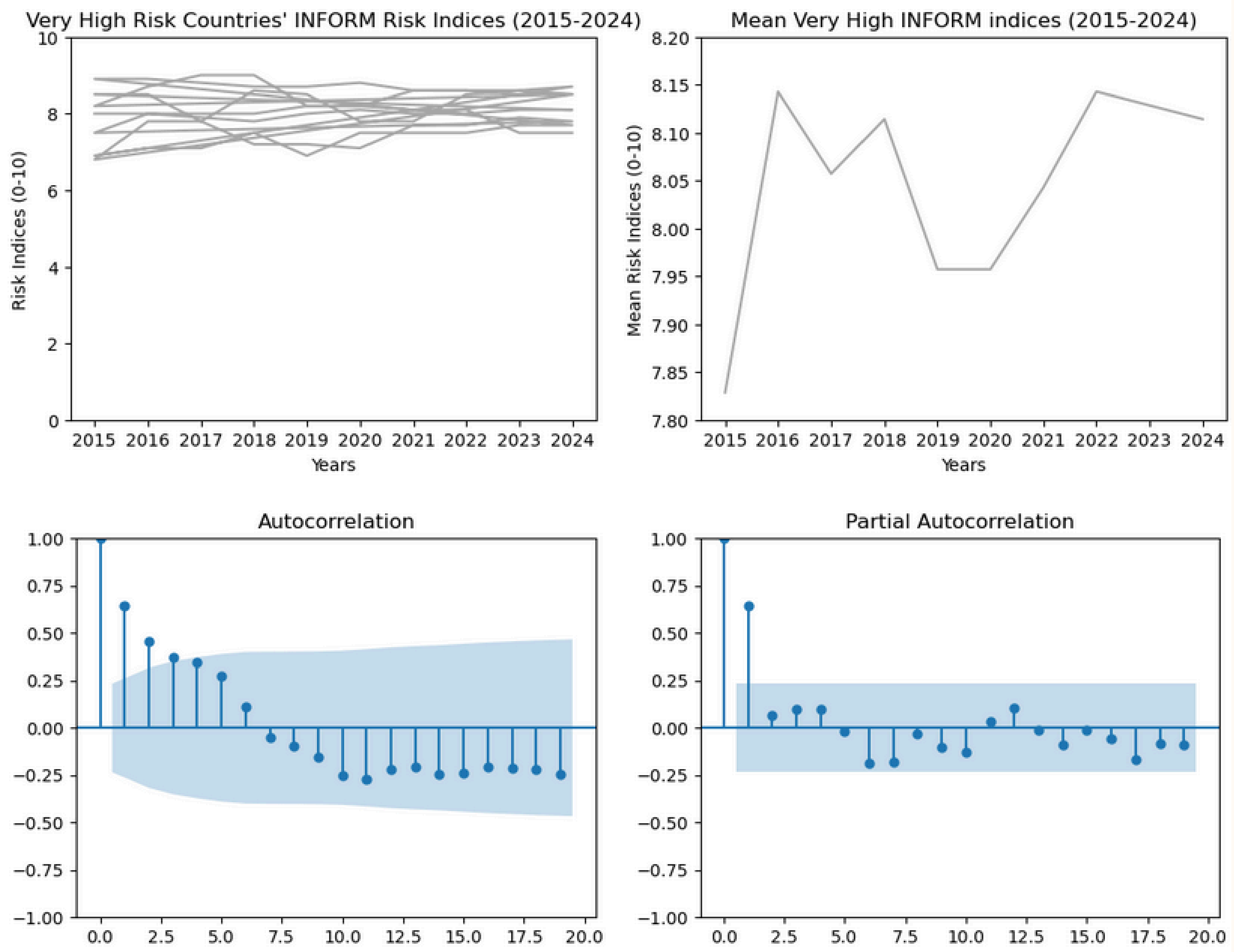


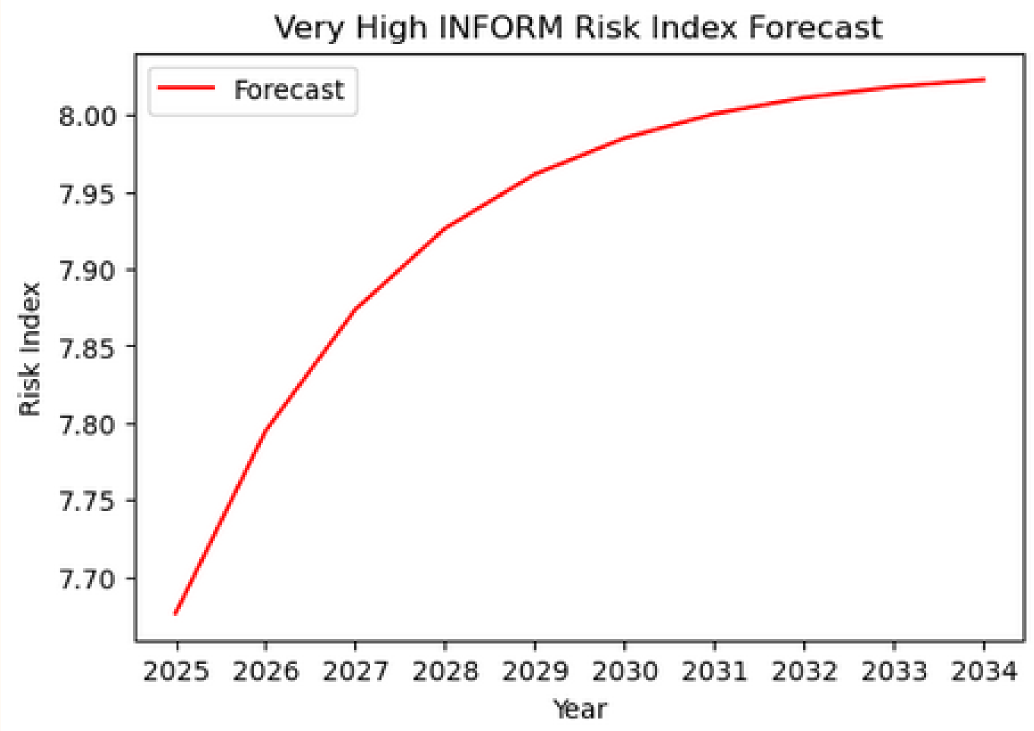

```
Mean Squared Error (MSE): 0.5623539349422726
Mean Absolute Error (MAE): 0.6229084757735233
Root Mean Squared Error (RMSE): 0.7499026169725457
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:      INFORM Risk Index   No. Observations:                  310
Model:                 ARIMA(1, 1, 1)   Log Likelihood                -194.659
Date:                Mon, 08 Apr 2024   AIC                            395.318
Time:                        03:35:23   BIC                            406.518
Sample:                             0   HQIC                           399.795
                                - 310
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8215      0.047     17.302      0.000       0.728       0.915
ma.L1         -0.9888      0.013    -77.361      0.000      -1.014      -0.964
sigma2         0.2053      0.011     19.222      0.000       0.184       0.226
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):               872.08
Prob(Q):                              0.93   Prob(JB):                         0.00
Heteroskedasticity (H):               1.14   Skew:                            -0.87
Prob(H) (two-sided):                  0.49   Kurtosis:                        11.05
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

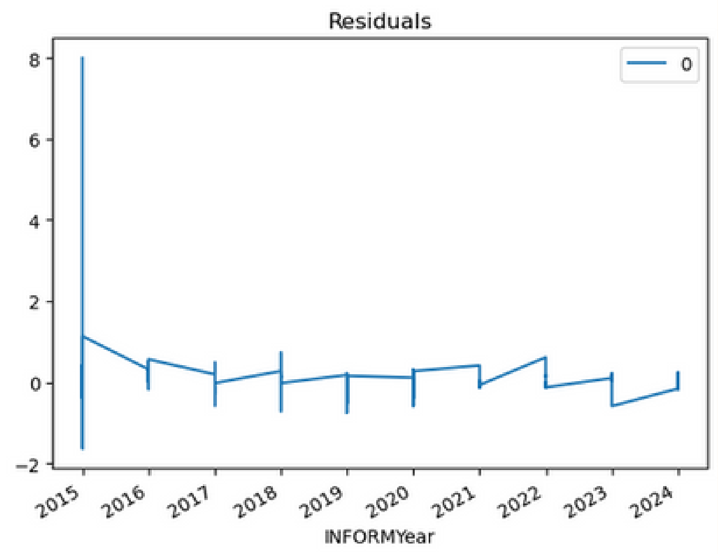
ARIMA Model to forecast Very High Risk Countries' INFORM Risk indices:

Back to Agenda






```
Mean Squared Error (MSE): 0.07576890740383165
Mean Absolute Error (MAE): 0.22655634355405924
Root Mean Squared Error (RMSE): 0.2752615254695644
```




```
                               SARIMAX Results
==============================================================================
Dep. Variable:      INFORM Risk Index   No. Observations:                   70
Model:                 ARIMA(1, 1, 1)   Log Likelihood                 -39.666
Date:                Mon, 08 Apr 2024   AIC                             85.331
Time:                        03:39:26   BIC                             92.034
Sample:                             0   HQIC                            87.990
                                 - 70
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.6683      0.127      5.274      0.000       0.420       0.917
ma.L1         -0.9992      2.783     -0.359      0.720      -6.455       4.456
sigma2         0.1779      0.502      0.355      0.723      -0.806       1.162
===================================================================================
Ljung-Box (L1) (Q):                   0.29   Jarque-Bera (JB):                72.85
Prob(Q):                              0.59   Prob(JB):                         0.00
Heteroskedasticity (H):               0.96   Skew:                            -1.29
Prob(H) (two-sided):                  0.93   Kurtosis:                         7.32
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Back to Agenda

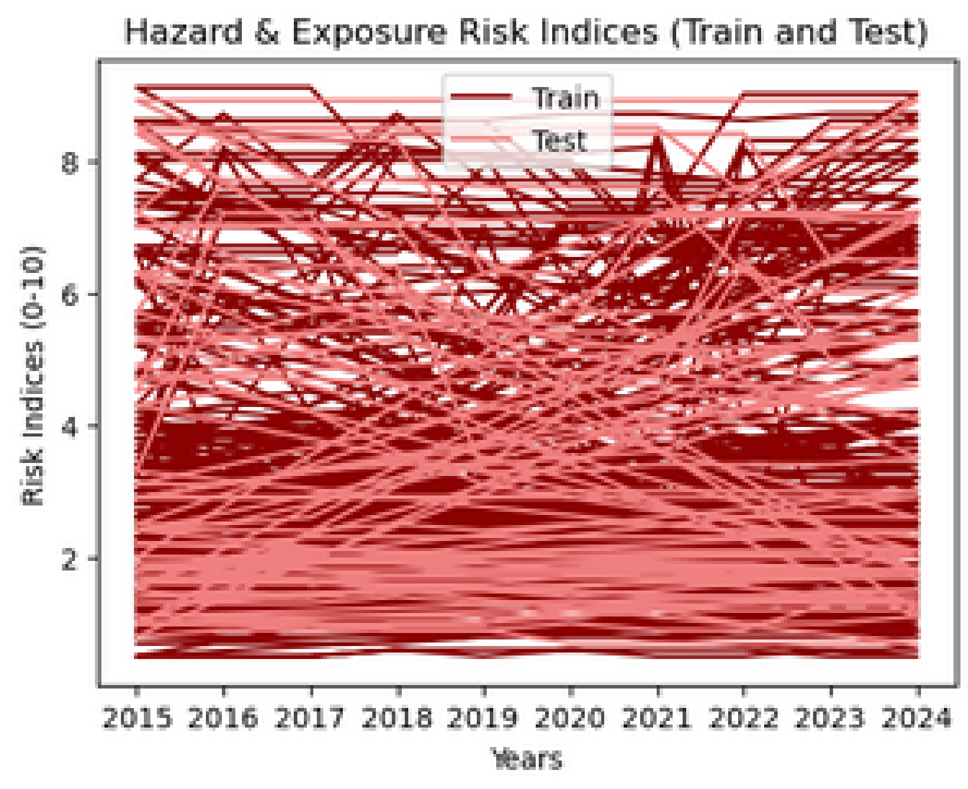
Project 3

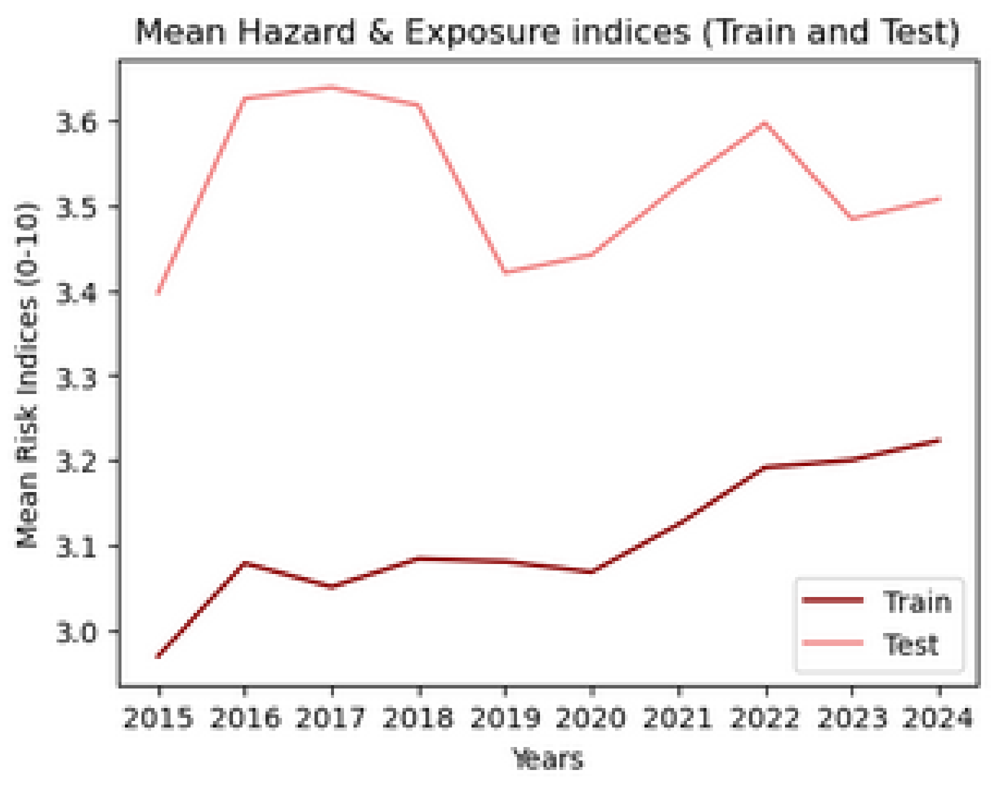
Predict & forecast how INFORM risk data changes over time

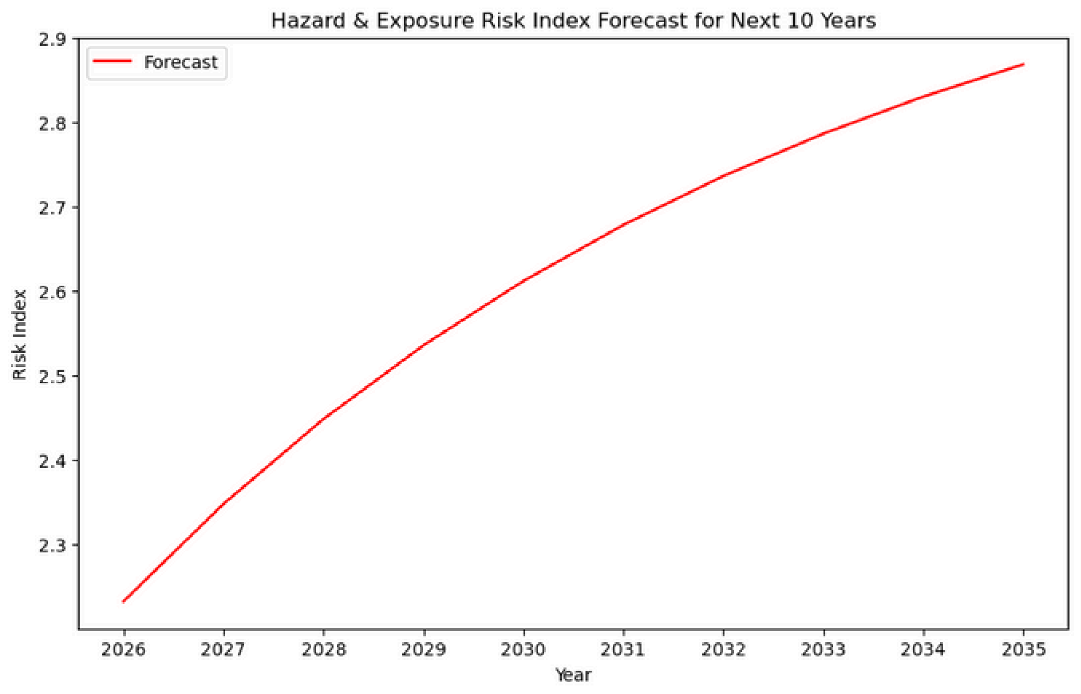
Hazard & Exposure Index

ARIMA Train - Test Model to forecast Hazard & Exposure Risk indices:

Back to Agenda





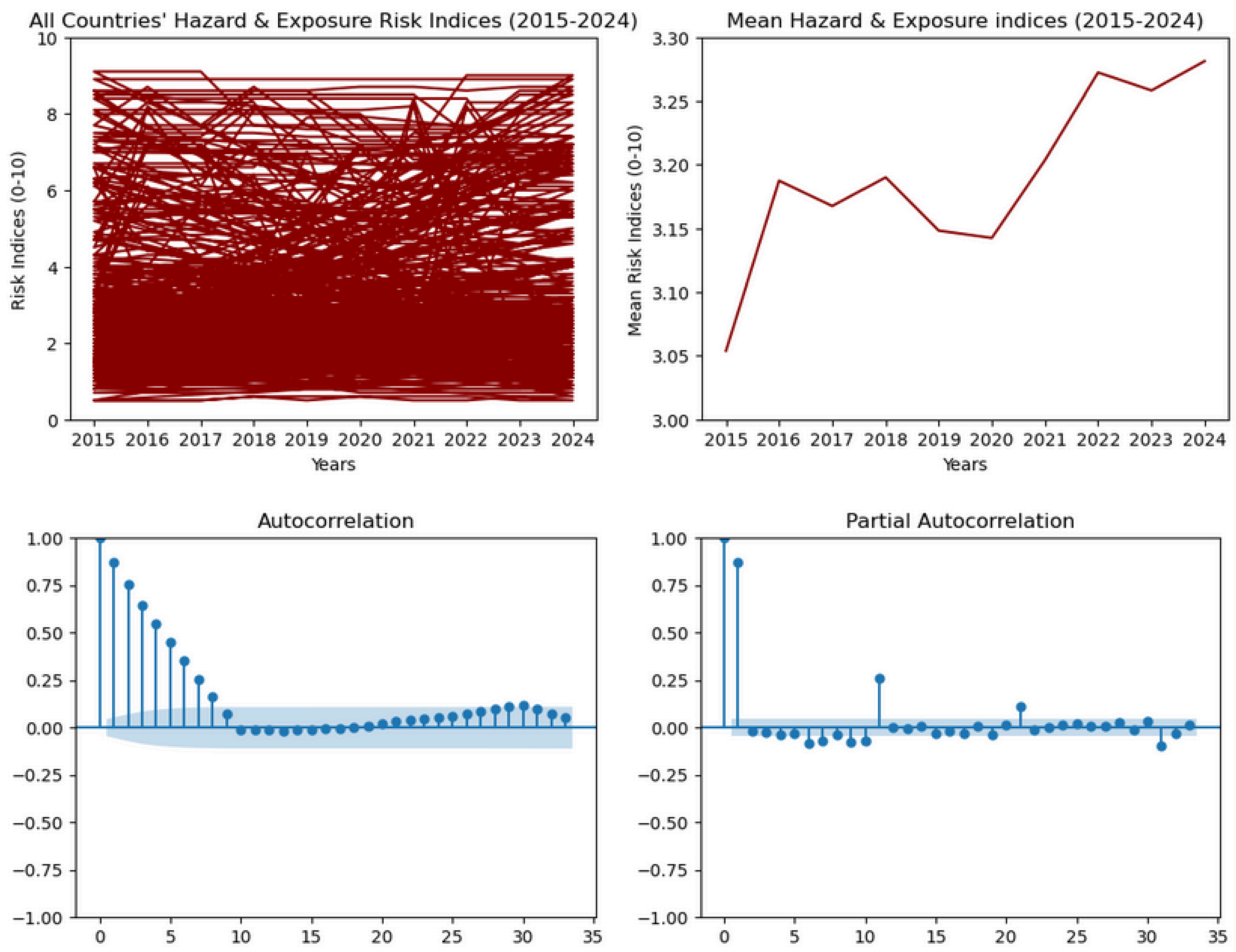


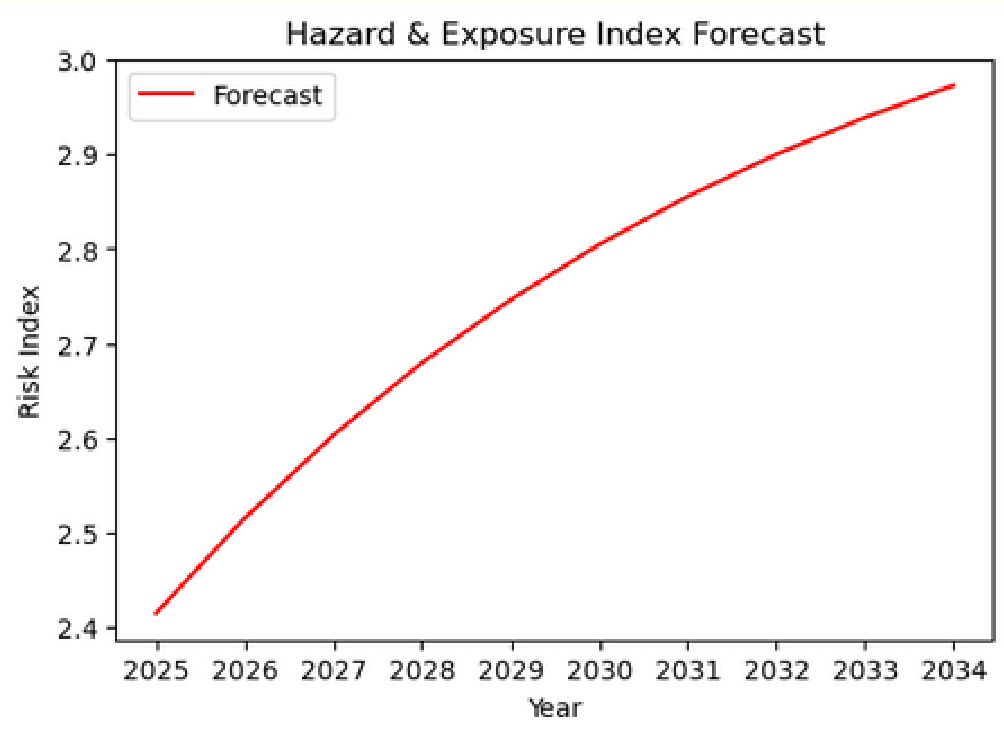

```
Mean Squared Error (MSE): 5.902127035081362
Mean Absolute Error (MAE): 1.9747482923871407
Root Mean Squared Error (RMSE): 2.4294293640855997
```

```
                                  SARIMAX Results
===========================================================================================
Dep. Variable:     Hazard & Exposure Index   No. Observations:                     1528
Model:                      ARIMA(1, 1, 1)   Log Likelihood                   -2198.746
Date:                     Mon, 08 Apr 2024   AIC                               4403.493
Time:                             22:08:47   BIC                               4419.486
Sample:                                  0   HQIC                              4409.445
                                    - 1528
Covariance Type:                       opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8709      0.013     69.268      0.000       0.846       0.896
ma.L1         -0.9999      0.047    -21.264      0.000      -1.092      -0.908
sigma2         1.0398      0.052     19.822      0.000       0.937       1.143
===========================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                 17099.64
Prob(Q):                              0.88   Prob(JB):                             0.00
Heteroskedasticity (H):               1.40   Skew:                                -0.12
Prob(H) (two-sided):                  0.00   Kurtosis:                            19.39
===========================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

ARIMA Model to forecast Hazard & Exposure Risk indices:

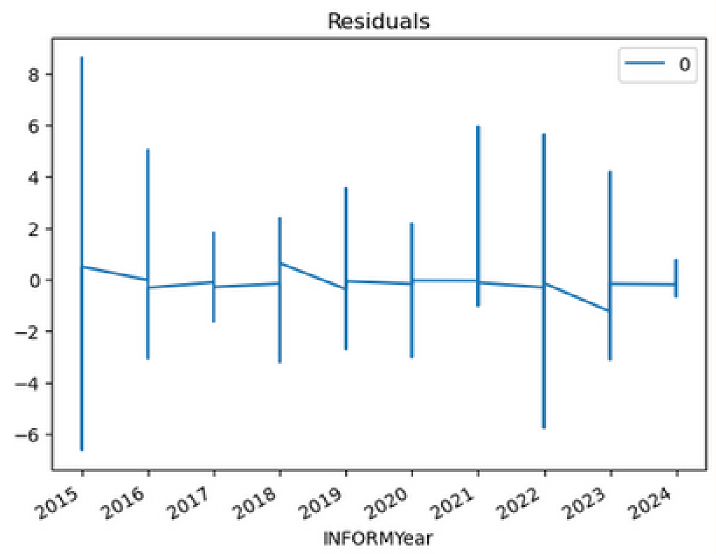
Back to Agenda






```
Mean Squared Error (MSE): 0.22468855088526024
Mean Absolute Error (MAE): 0.4530680413699869
Root Mean Squared Error (RMSE): 0.4740132391455541
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Hazard & Exposure Index   No. Observations:          1910
Model:                      ARIMA(1, 1, 1)   Log Likelihood         -2797.204
Date:                     Mon, 08 Apr 2024   AIC                     5600.407
Time:                             02:59:50   BIC                     5617.070
Sample:                                  0   HQIC                    5606.540
                                    - 1910
Covariance Type:                       opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8733      0.012     70.885      0.000       0.849       0.897
ma.L1         -1.0000      0.142     -7.022      0.000      -1.279      -0.721
sigma2         1.0943      0.153      7.160      0.000       0.795       1.394
===================================================================================
Ljung-Box (L1) (Q):                   0.24   Jarque-Bera (JB):             20759.99
Prob(Q):                              0.63   Prob(JB):                         0.00
Heteroskedasticity (H):               1.42   Skew:                             0.05
Prob(H) (two-sided):                  0.00   Kurtosis:                        19.16
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```



Back to Agenda

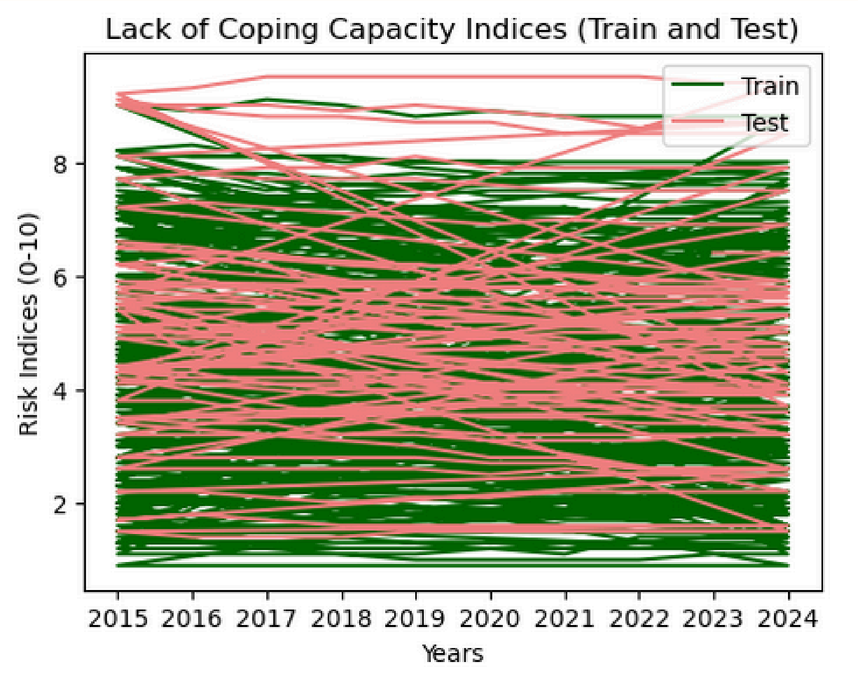
Project 3

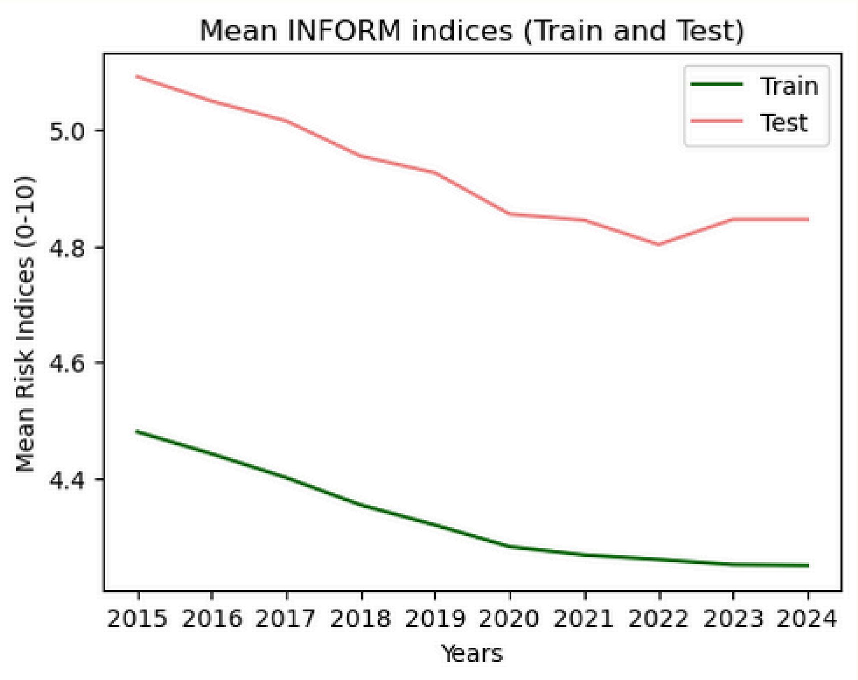
Predict & forecast how INFORM risk data changes over time

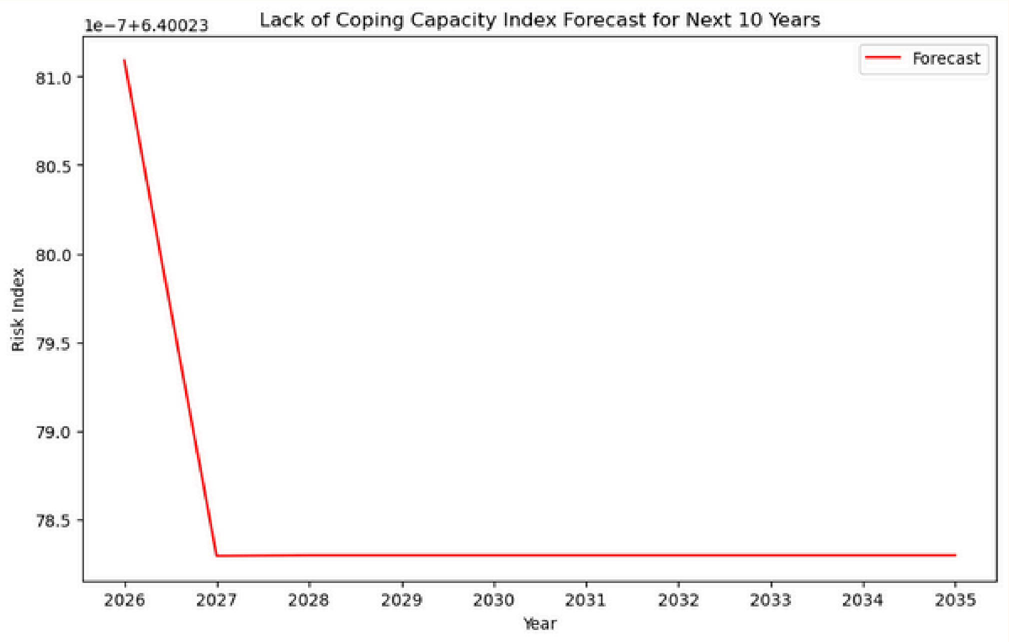
Lack of Coping Capacity Index

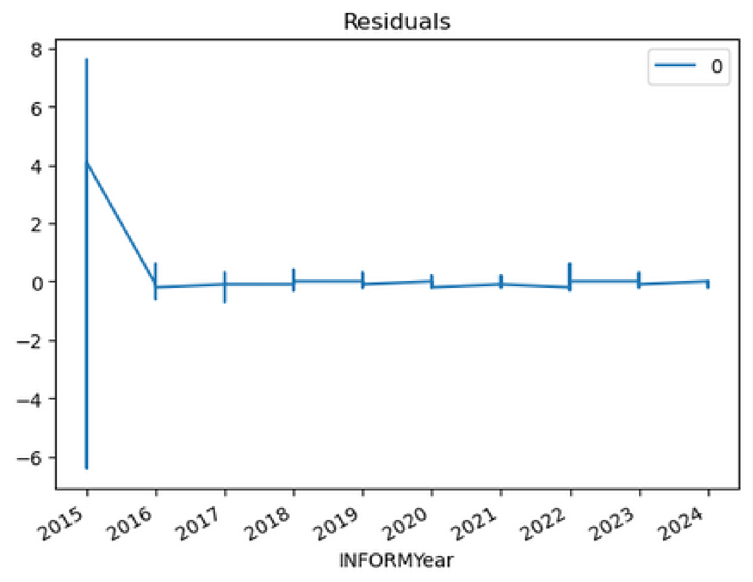
ARIMA Train-Test Model to forecast Lack of Coping Capacity Risk indices:

Back to Agenda







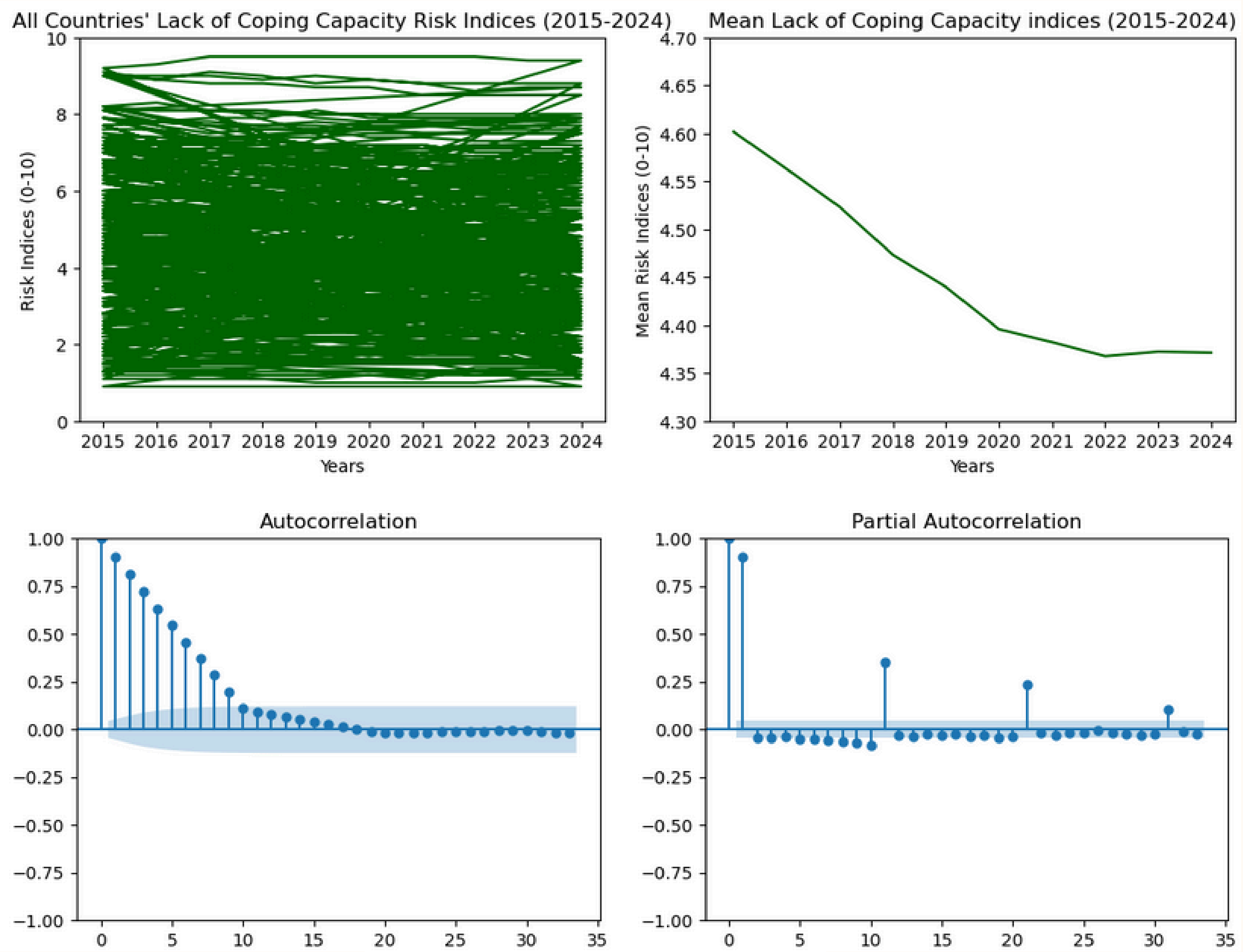


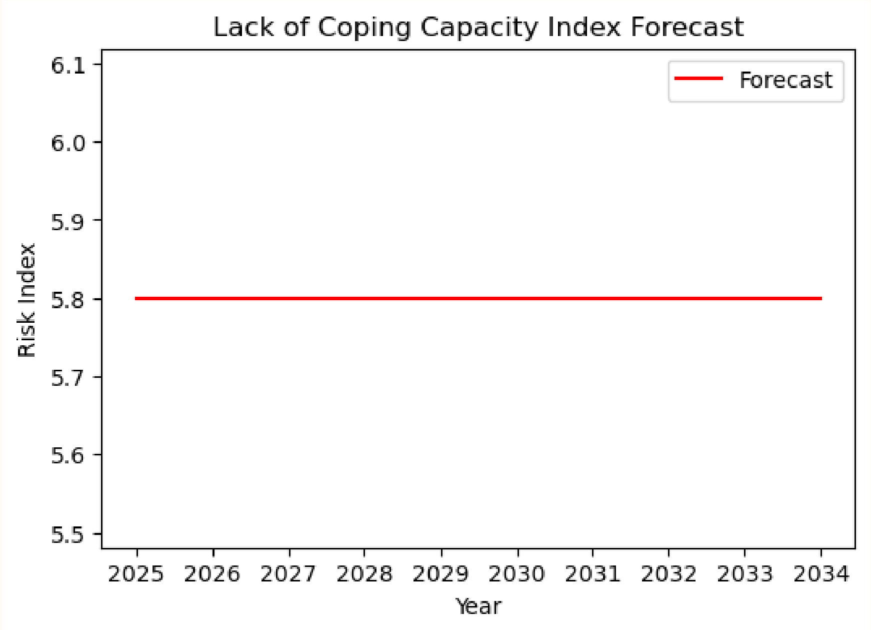

```
Mean Squared Error (MSE): 5.79462928977949
Mean Absolute Error (MAE): 2.064557289776779
Root Mean Squared Error (RMSE): 2.4072036244945068
```

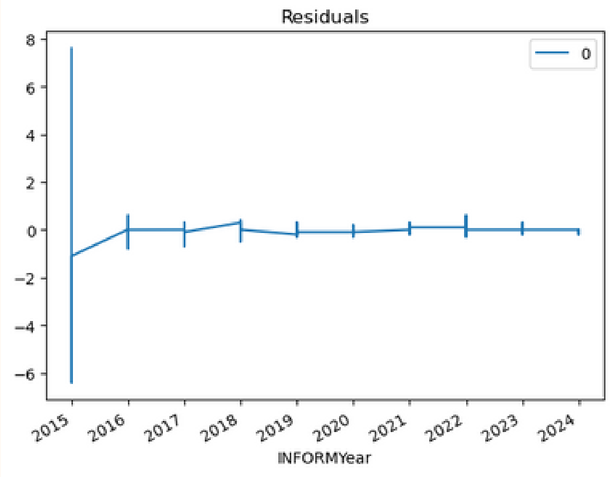
```
                                    SARIMAX Results
==========================================================================================
Dep. Variable:      Lack of Coping Capacity Index   No. Observations:                1528
Model:                             ARIMA(1, 1, 1)   Log Likelihood              -1902.988
Date:                            Mon, 08 Apr 2024   AIC                          3811.976
Time:                                    22:28:02   BIC                          3827.969
Sample:                                         0   HQIC                         3817.928
                                           - 1528
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.0012     50.073  -2.35e-05      1.000     -98.142      98.139
ma.L1         -0.0012     50.072  -2.41e-05      1.000     -98.140      98.138
sigma2         0.7079      0.008     93.769      0.000       0.693       0.723
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):             28310.06
Prob(Q):                              1.00   Prob(JB):                         0.00
Heteroskedasticity (H):               1.33   Skew:                            -0.08
Prob(H) (two-sided):                  0.00   Kurtosis:                        24.09
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

ARIMA Model to forecast Lack of Coping Capacity Risk indices:

Back to Agenda








```
Mean Squared Error (MSE): 0.0010000000000000106
Mean Absolute Error (MAE): 0.010000000000000054
Root Mean Squared Error (RMSE): 0.031622776601683965
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Lack of Coping Capacity Index   No. Observations:      1910
Model:                            ARIMA(1, 1, 1)   Log Likelihood    -2349.559
Date:                           Mon, 08 Apr 2024   AIC                4705.118
Time:                                   03:06:58   BIC                4721.781
Sample:                                        0   HQIC               4711.251
                                          - 1910
Covariance Type:                             opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.0014     37.000  -3.68e-05      1.000     -72.520      72.517
ma.L1         -0.0016     36.999  -4.33e-05      1.000     -72.518      72.515
sigma2         0.6864      0.007    105.031      0.000       0.674       0.699
==================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):             35639.61
Prob(Q):                              1.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.95   Skew:                             0.14
Prob(H) (two-sided):                  0.49   Kurtosis:                        24.17
==================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Back to Agenda

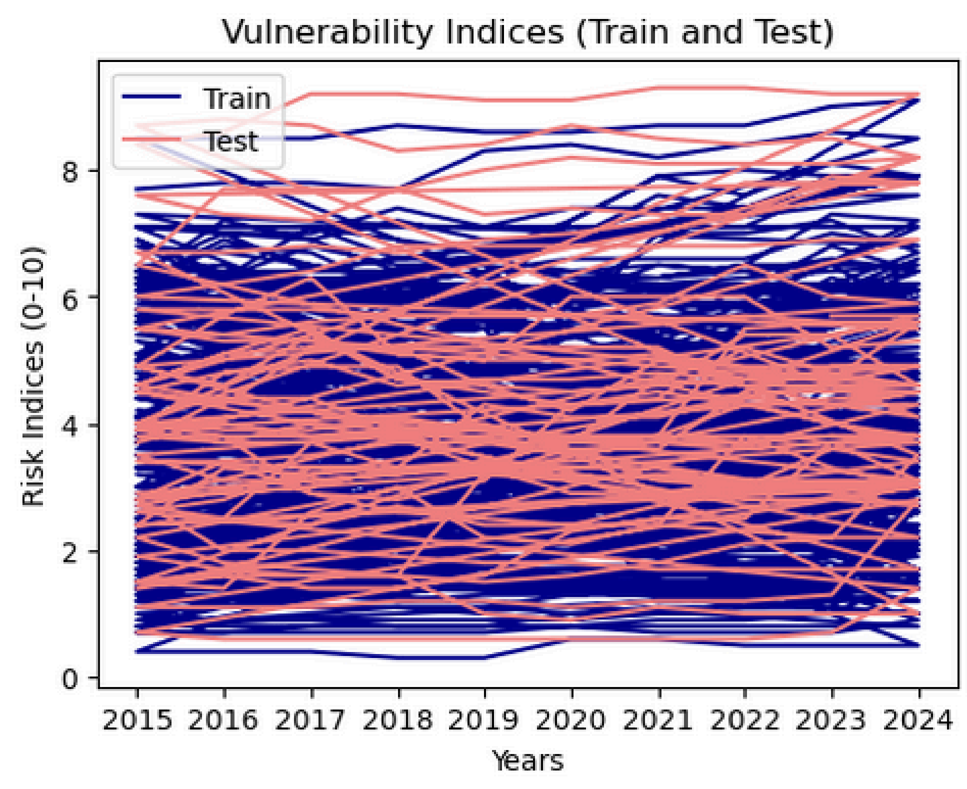
Project 3

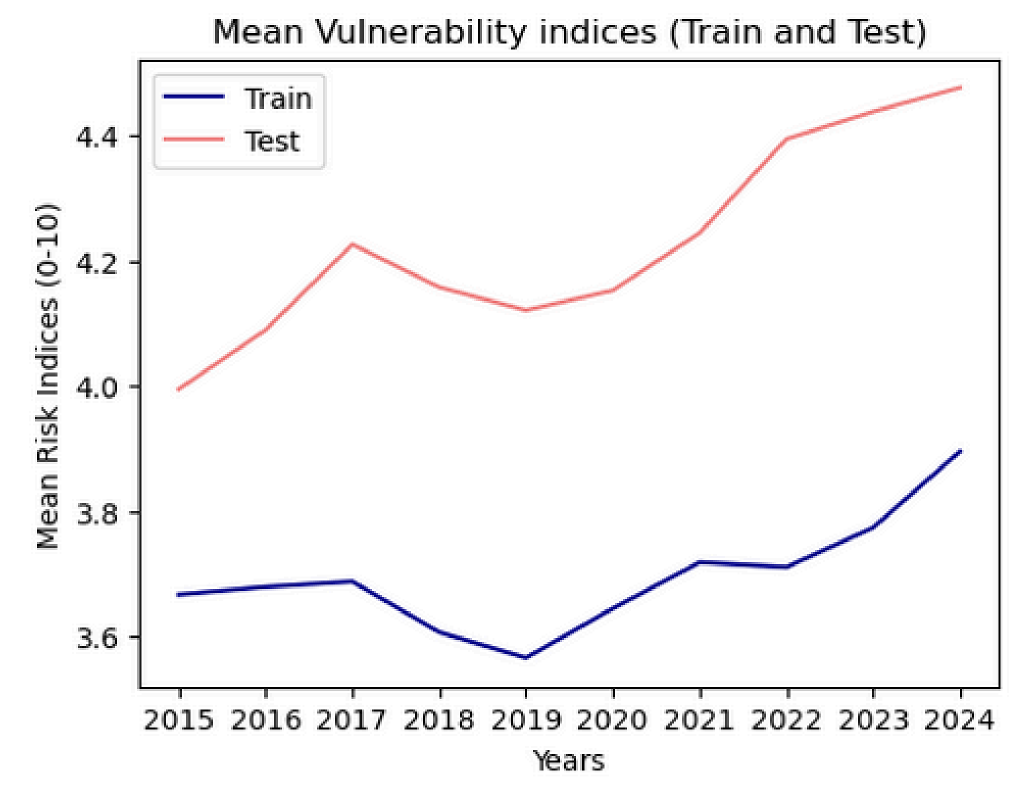
Vulnerability Index

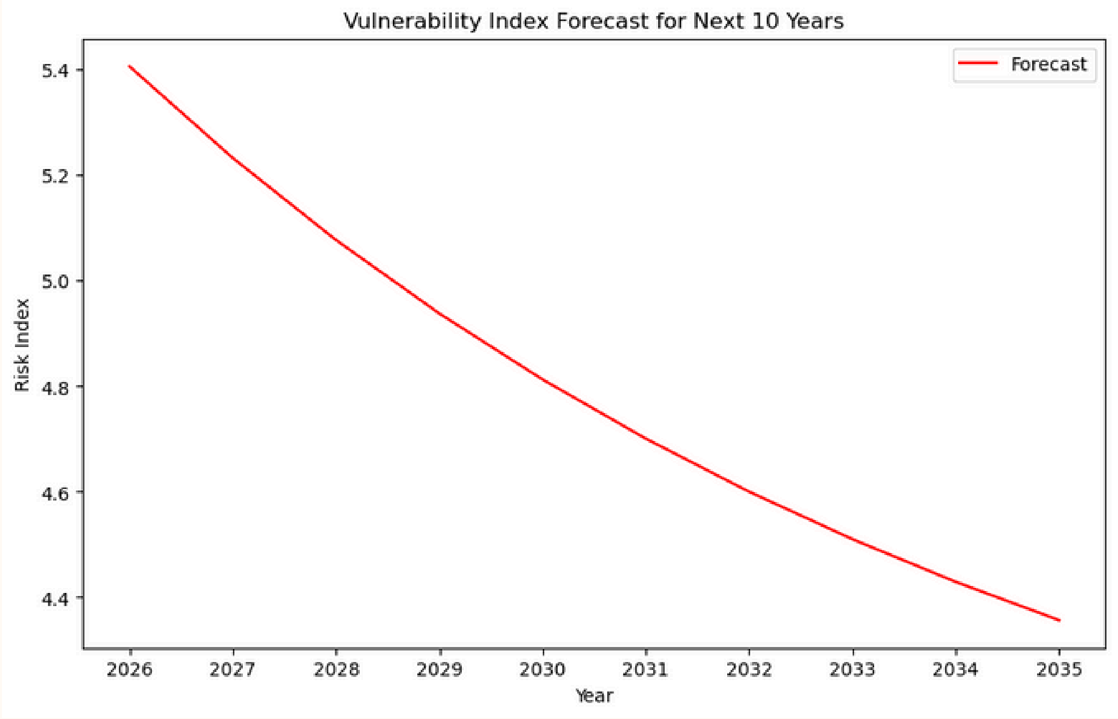
Predict & forecast how INFORM risk data changes over time

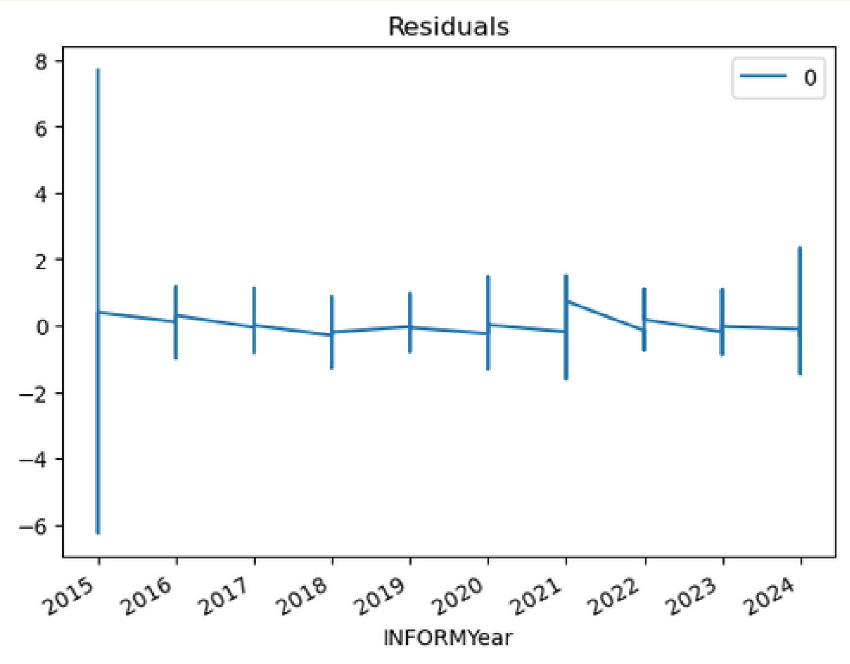
ARIMA Train - Test Model to forecast Vulnerability Risk indices:

Back to Agenda







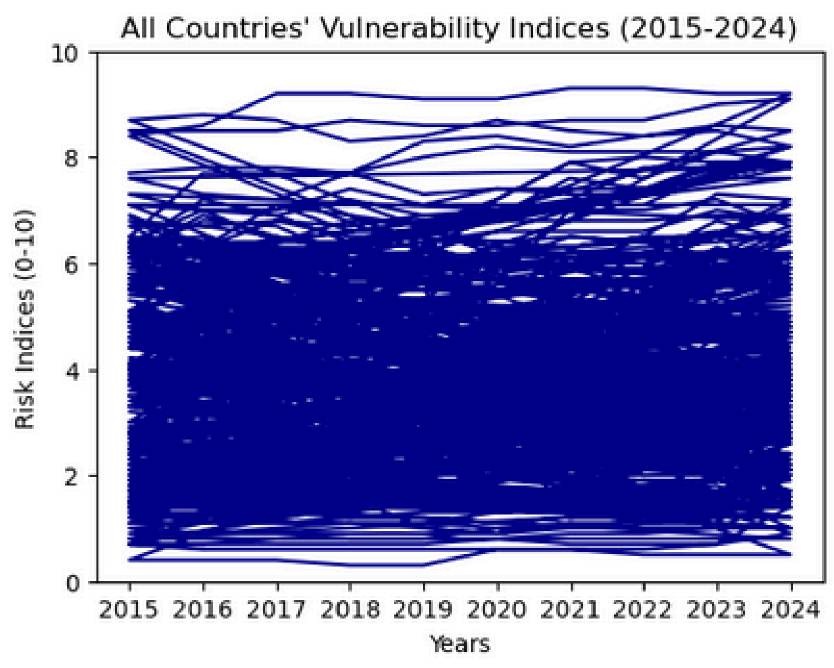


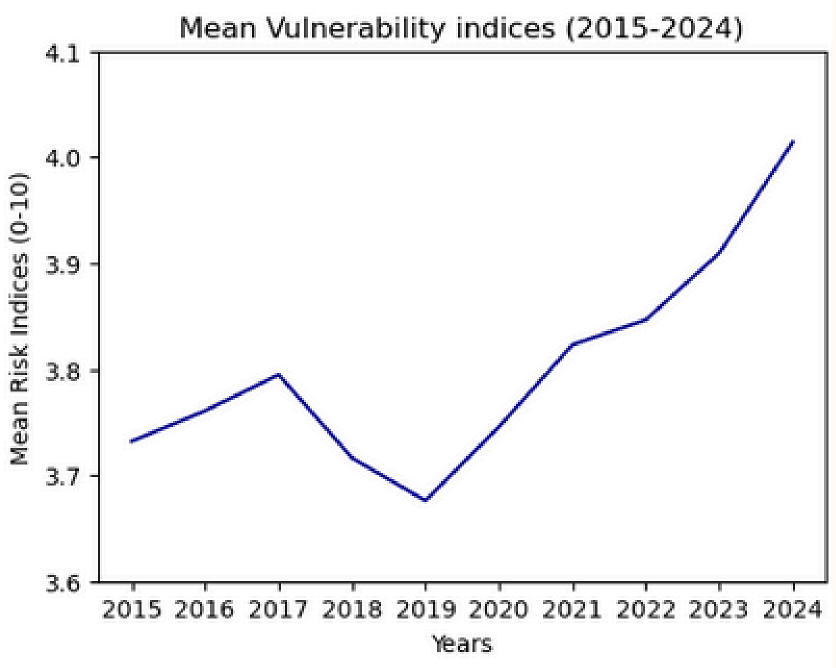

```
Mean Squared Error (MSE): 4.305888459231918
Mean Absolute Error (MAE): 1.5919862293228235
Root Mean Squared Error (RMSE): 2.0750634831811574
```

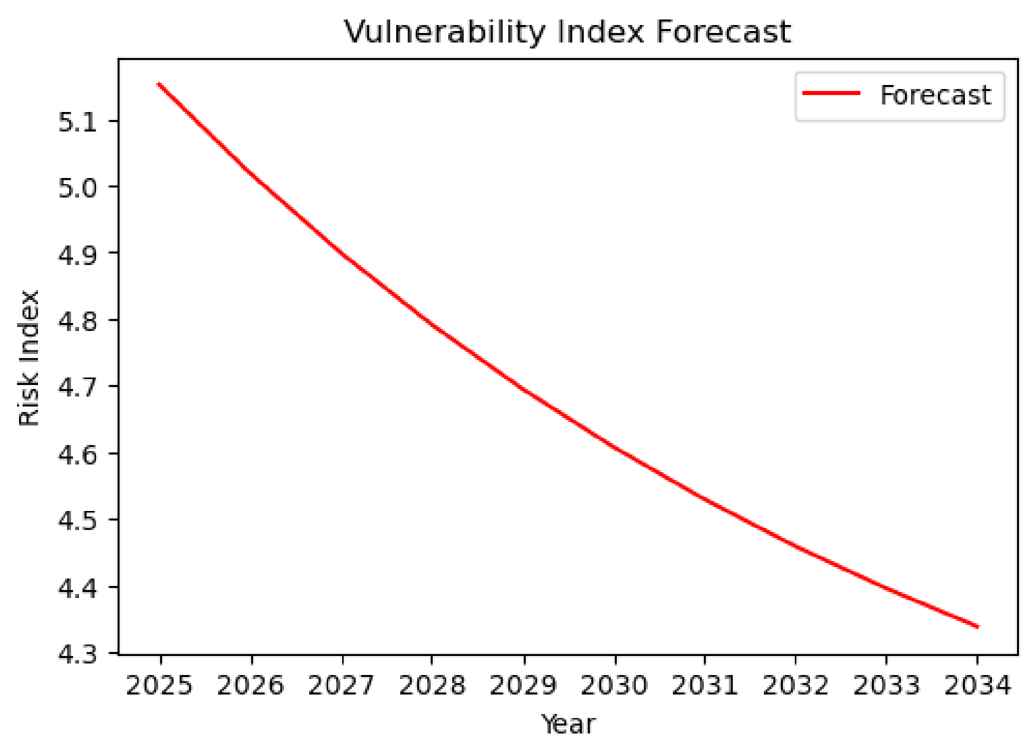
```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Vulnerability Index   No. Observations:                 1528
Model:                  ARIMA(1, 1, 1)   Log Likelihood               -1897.502
Date:                 Mon, 08 Apr 2024   AIC                           3801.003
Time:                         22:37:46   BIC                           3816.996
Sample:                              0   HQIC                          3806.956
                                - 1528
Covariance Type:                   opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8967      0.016     54.902      0.000       0.865       0.929
ma.L1         -1.0000      0.186     -5.373      0.000      -1.365      -0.635
sigma2         0.7008      0.133      5.264      0.000       0.440       0.962
===================================================================================
Ljung-Box (L1) (Q):                   4.55   Jarque-Bera (JB):             13727.54
Prob(Q):                              0.03   Prob(JB):                         0.00
Heteroskedasticity (H):               1.45   Skew:                            -0.42
Prob(H) (two-sided):                  0.00   Kurtosis:                        17.66
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

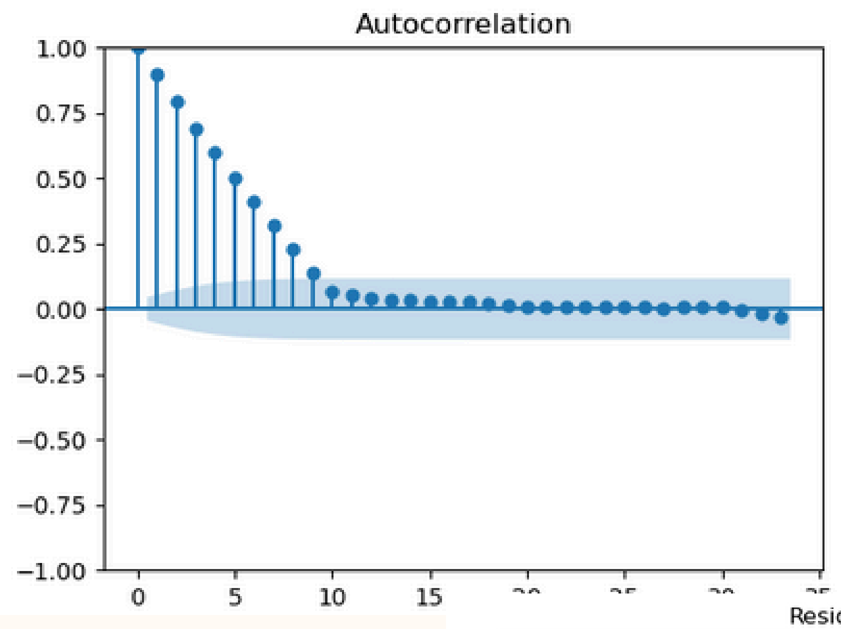
ARIMA Model to forecast Vulnerability Risk indices:

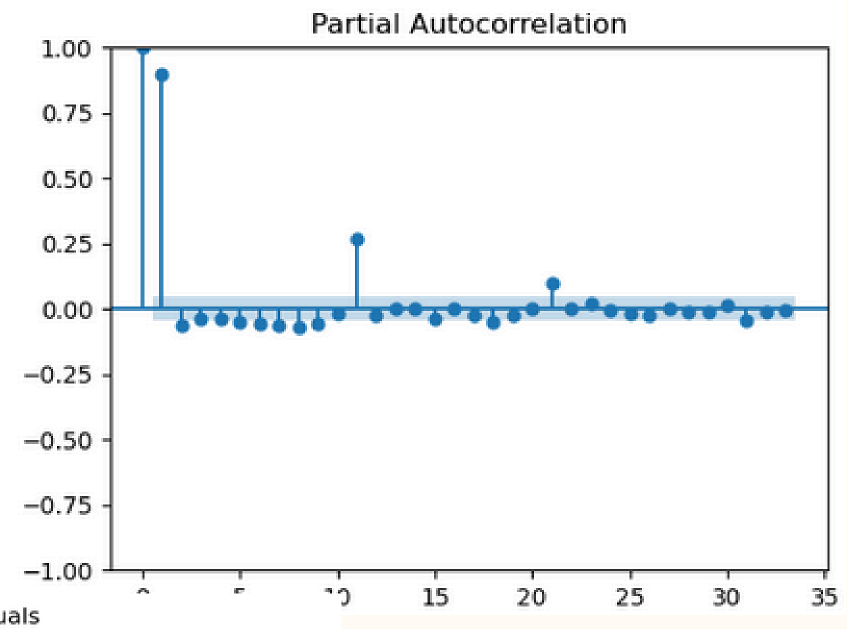
Back to Agenda












```
Mean Squared Error (MSE): 1.072450705808254
Mean Absolute Error (MAE): 0.9116852017083055
Root Mean Squared Error (RMSE): 1.035591959126882
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Vulnerability Index   No. Observations:                 1910
Model:                  ARIMA(1, 1, 1)   Log Likelihood               -2385.124
Date:                 Mon, 08 Apr 2024   AIC                           4776.247
Time:                         03:11:51   BIC                           4792.910
Sample:                              0   HQIC                          4782.380
                                - 1910
Covariance Type:                   opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8996      0.014     64.481      0.000       0.872       0.927
ma.L1         -1.0000      0.169     -5.912      0.000      -1.331      -0.668
sigma2         0.7107      0.119      5.969      0.000       0.477       0.944
===================================================================================
Ljung-Box (L1) (Q):                   5.45   Jarque-Bera (JB):             16503.83
Prob(Q):                              0.02   Prob(JB):                         0.00
Heteroskedasticity (H):               1.41   Skew:                            -0.27
Prob(H) (two-sided):                  0.00   Kurtosis:                        17.39
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Back to Agenda

Summary

Summary Project# 3

Back to Agenda

All Countries:

- Test-train didn't improve the fit or forecast output
- Using full dataset had better model fit & performance

INFORM Index

- Limited/poor model goodness of fit
- Good model performance

Hazard & Exposure

- Limited/poor model goodness of fit
- Good model performance

Lack of Coping Capacity

- Limited/poor model goodness of fit
- Good model performance

Vulnerability

- Limited/poor model goodness of fit
- Limited/poor model performance

Summary Project# 3

Back to Agenda

For Risk Groups:

INFORM Index

- Good model goodness of fit for V.low, Low, Medium Risk Countries
- Good model performance for V.Low, Low & V.High Risk Countries

Hazard & Exposure

- Good model goodness of fit for V.low & Low Risk Countries
- Good model performance for V.High Risk Countries

Lack of Coping Capacity

- Good model goodness of fit for V.low, Low, High & V.High Risk Countries
- Good model performance for V.High Risk Countries

Vulnerability

- Poor model goodness of fit
- Good model performance for V.High Risk Countries

Back to Agenda

Conclusions

Conclusions Project 1 + 2

- The INFORM Risk index **can** be used to predict humanitarian crises, to a certain extent
- Considering the scale of the data, sometimes simplicity is best
  For example- logisitic model performed better than decision tree
- XGBoost was best performing model overall
- There is a lot of room for improvement

Conclusions Project 3

Back to Agenda

ARIMA model can provide predictions of future Risk Indices

Was ARIMA model most appropriate?
Would simpler **Autoregression** or **Moving Average** be sufficient for prediction? (Autocorrelation & Partial Autocorrelation Plots)

Associations with Core Indicators suggest multivariate model may improve predictions
Useful as indices data on a limited scale & lacks granularity for time series forecasting, ie 10 data points per country over 10 year period

Could more training & testing for risk groups better fit the model improve predictions?

Could **AutoARIMA** be equal or superior to the ARIMA model?

Next Steps Project 1 + 2

- Determine why the linear regression performed so badly on Project 2
- Take into account the TYPE of crisis
- Look more into individual features
- Finding the best hyperparameters for XGBoost

Next Steps Project 3

Back to Agenda

Further EDA between specific **INFORM Indices vs Core Indicators** for use in **multivariate models**

Explore whether **simpler models**, eg **Autoregression** or **Moving Average** would improve on ARIMA forcasting

Explore & identify **more granular data for time series forecasting**
Increase data points per month
For specific country/ group of countries

Determine why the linear regression performed so badly on Project 2

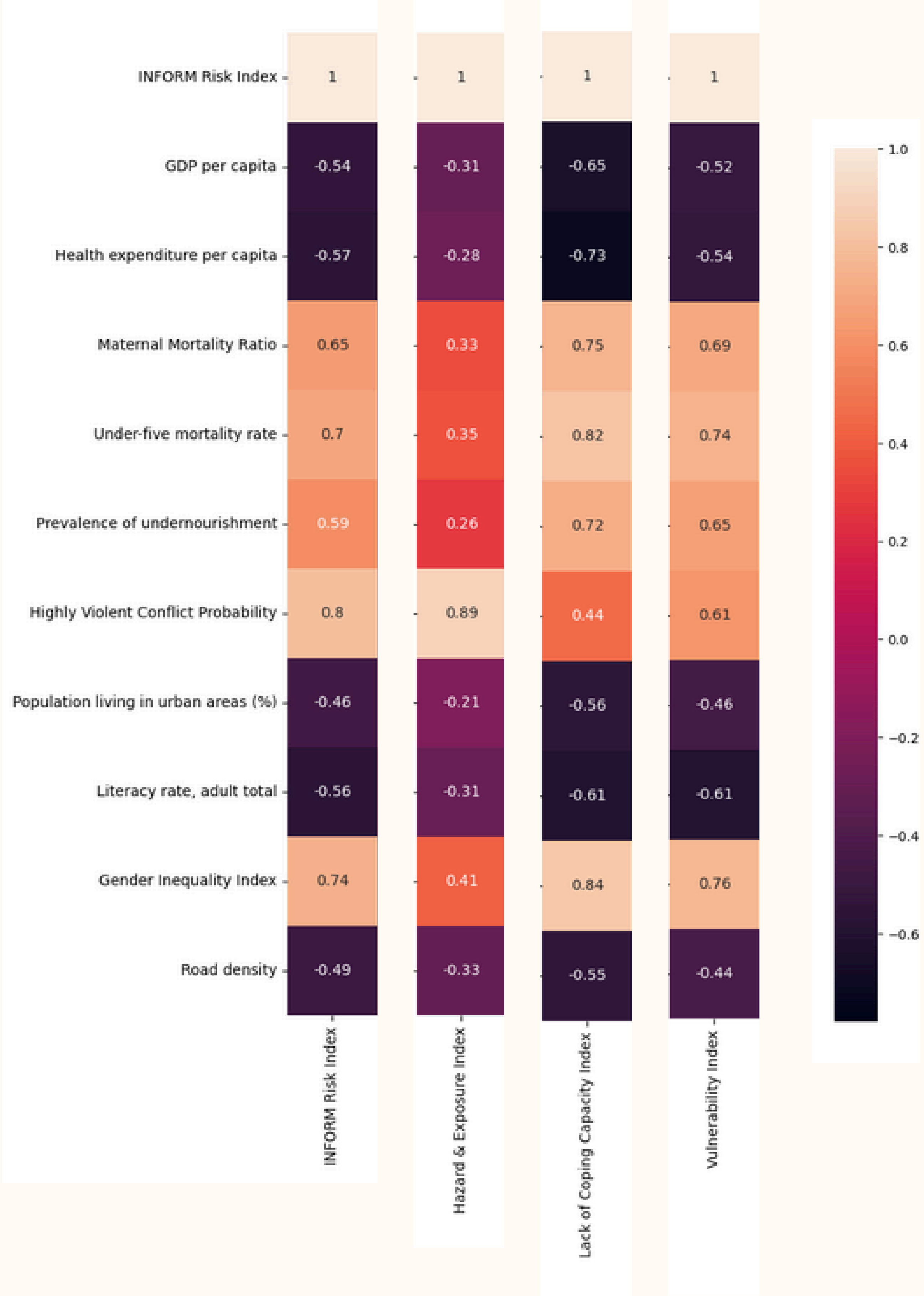
Thank you!

Back to Agenda

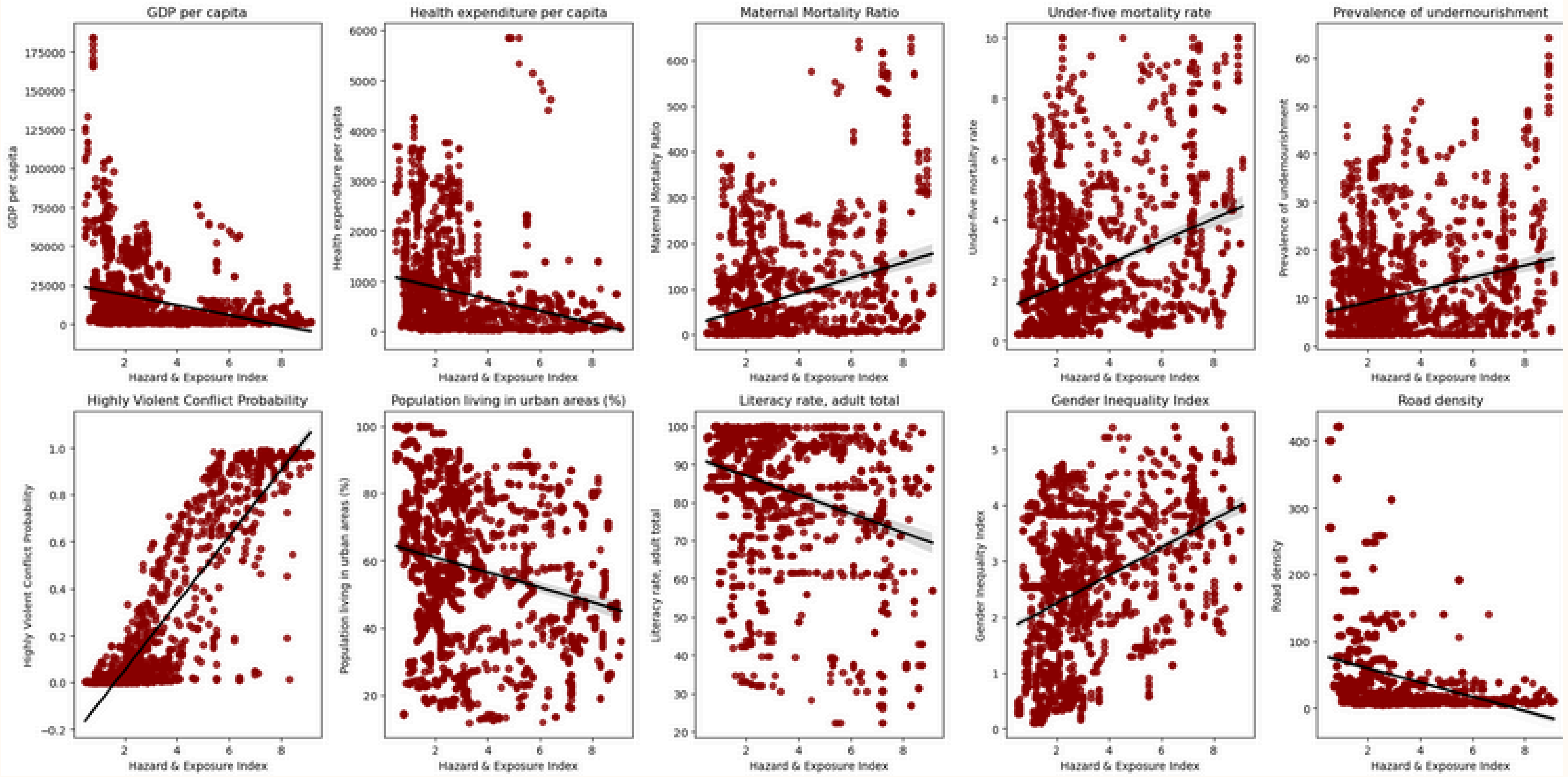
Additional Slides

EDA

Back to Agenda

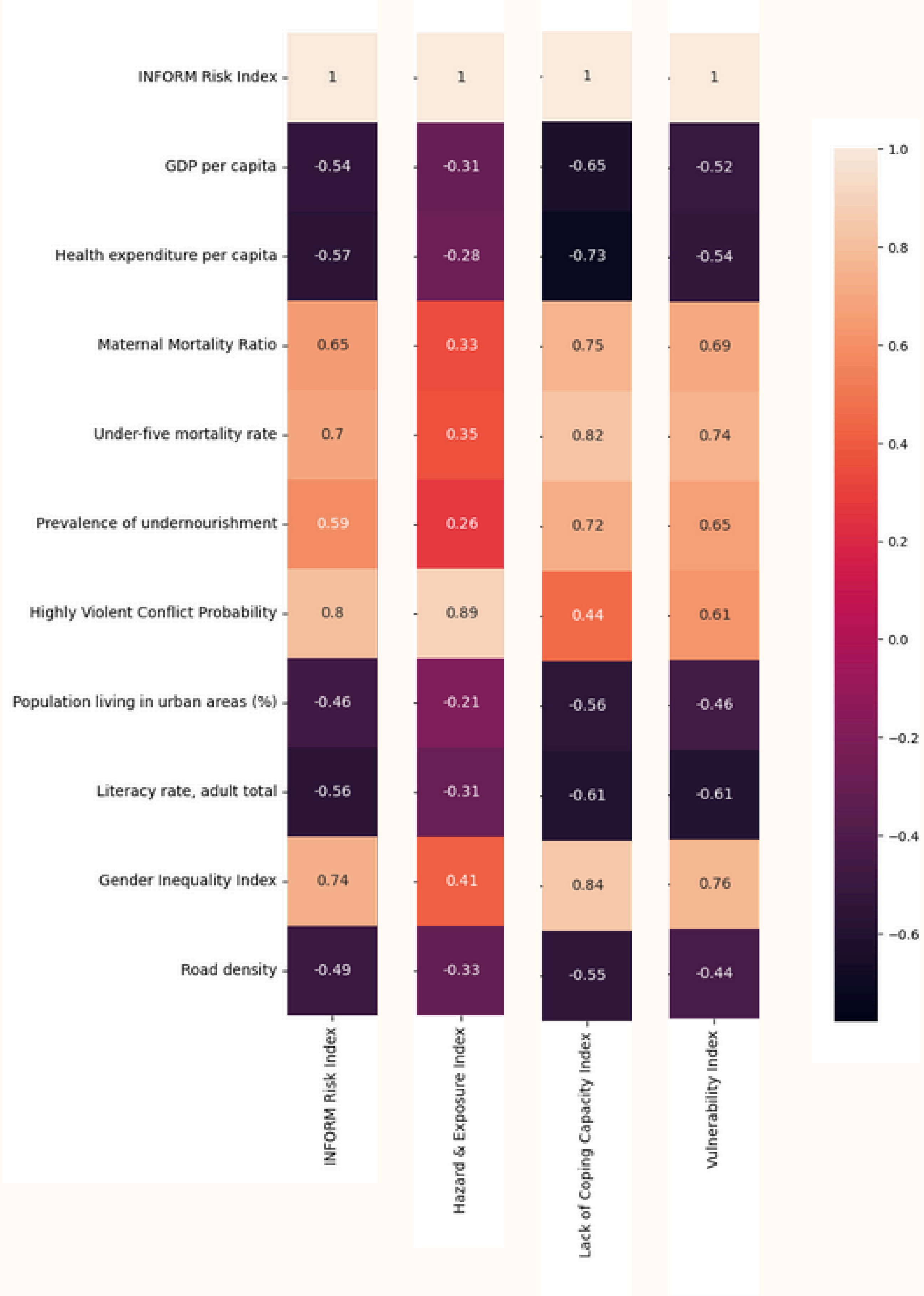


Correlations: Hazard & Exposure Index vs Core Indicators

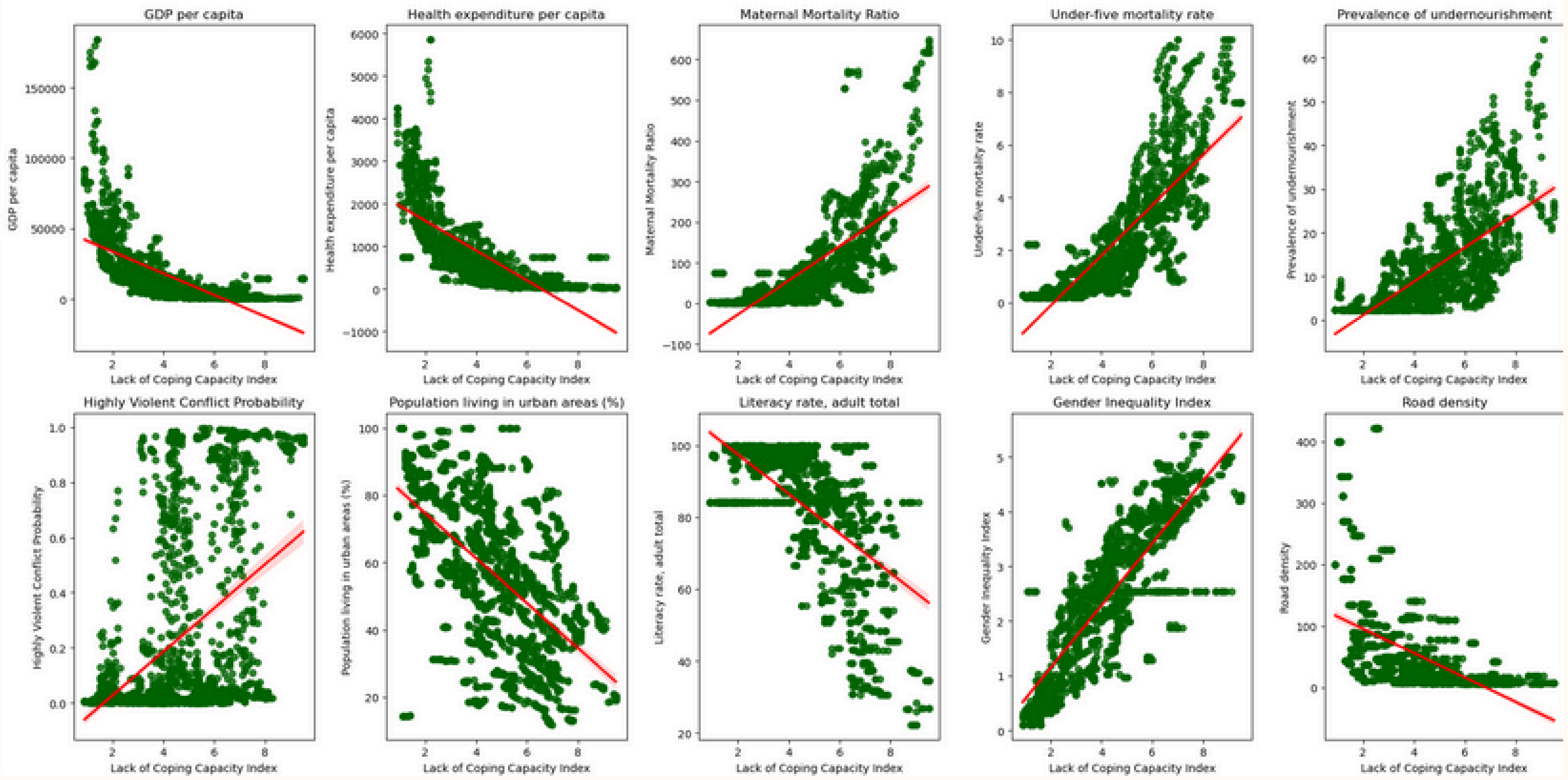


EDA

Back to Agenda

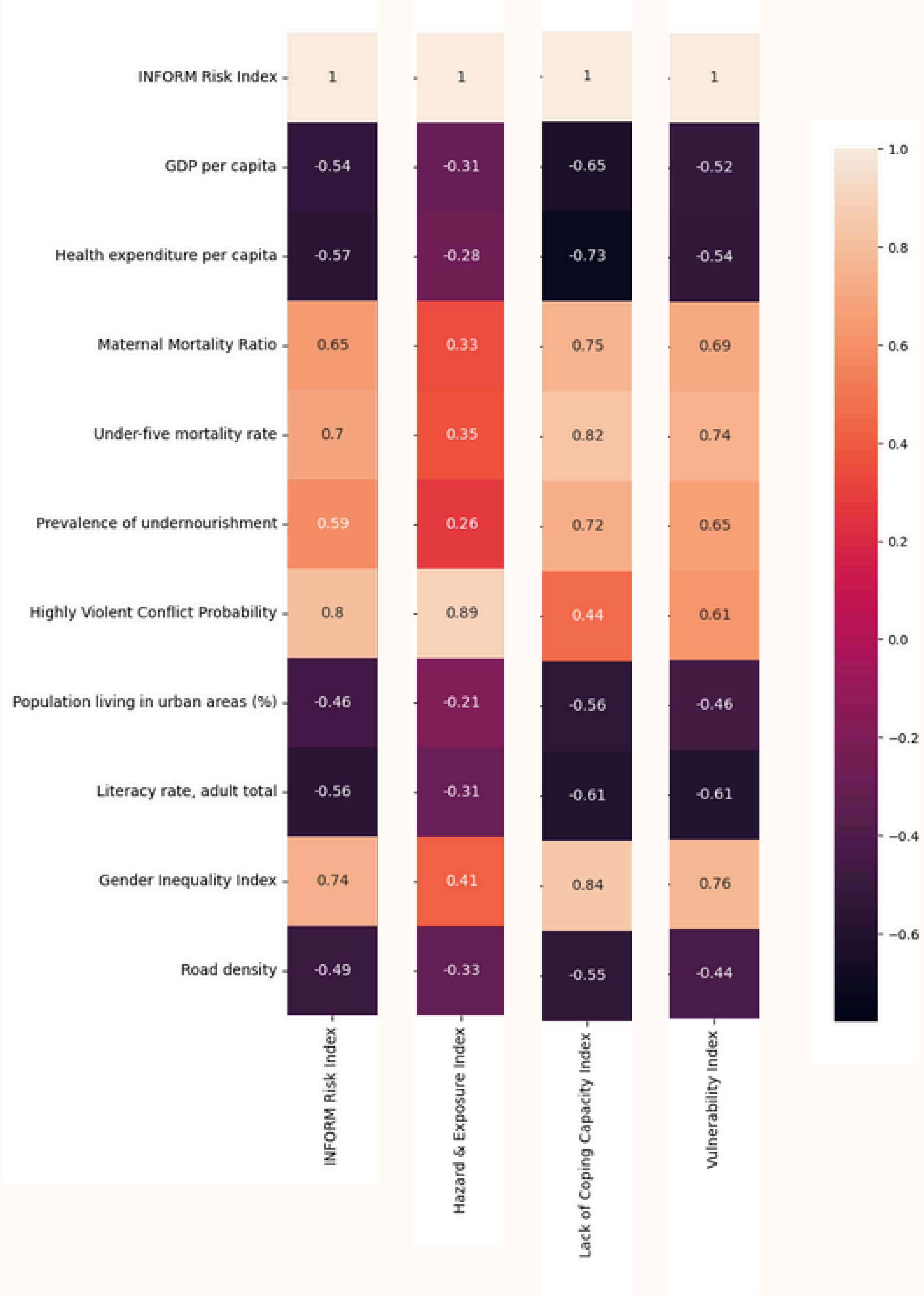


Correlations: Lack of Coping Capacity Index vs Core Indicators

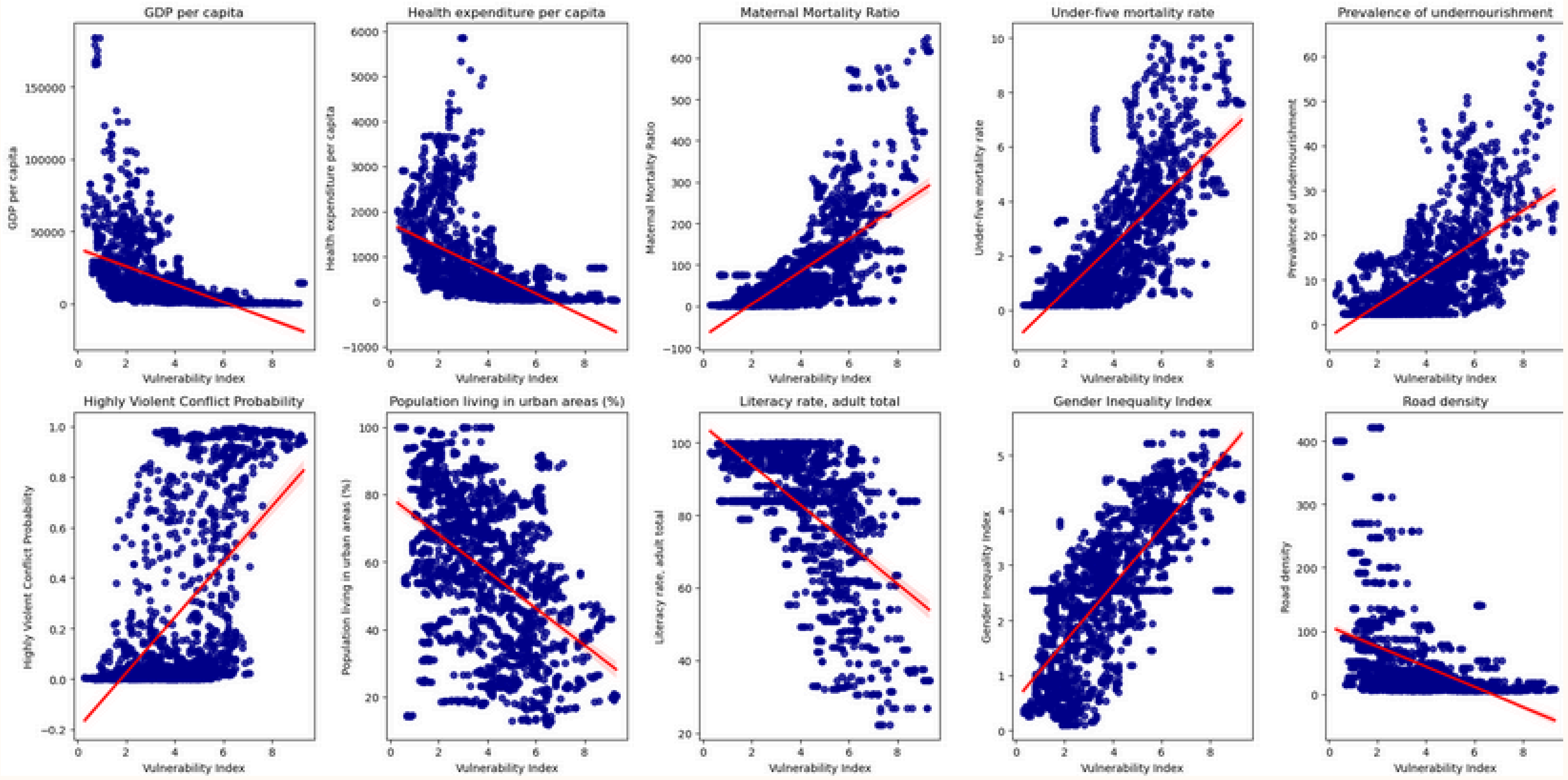


EDA

Back to Agenda



Correlations: Vulnerability Index vs Core Indicators



Back to Agenda

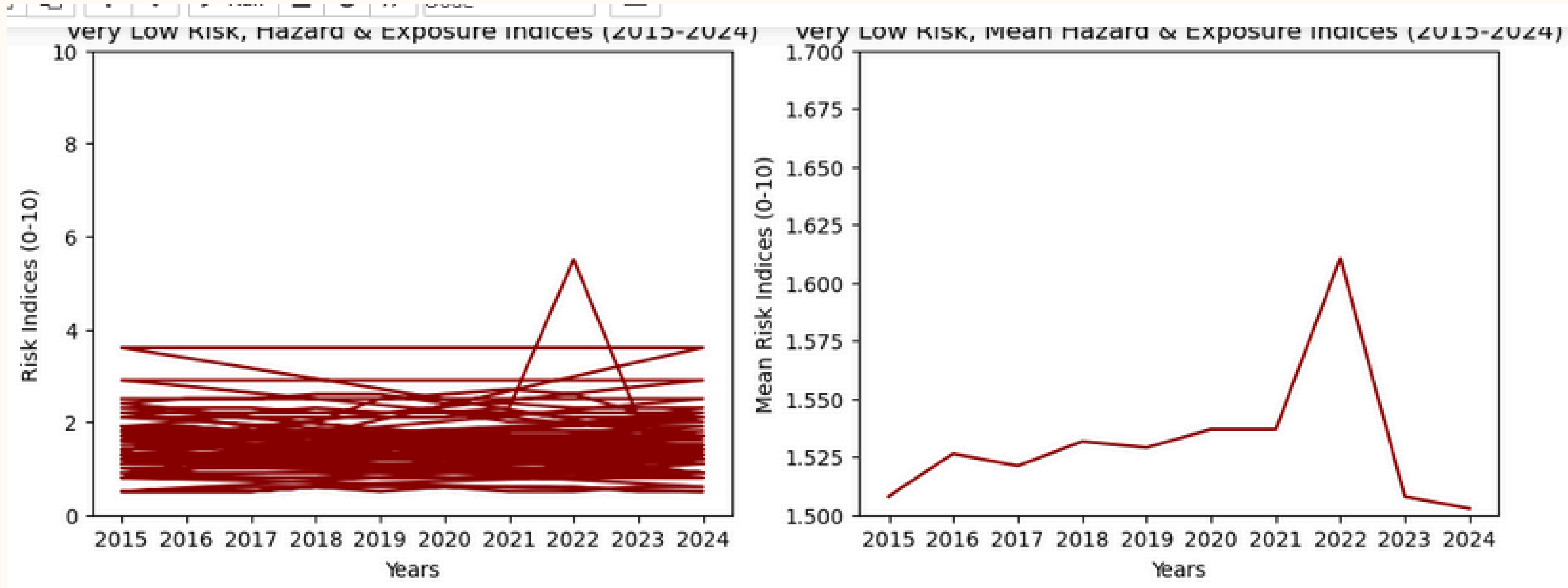
Project 3

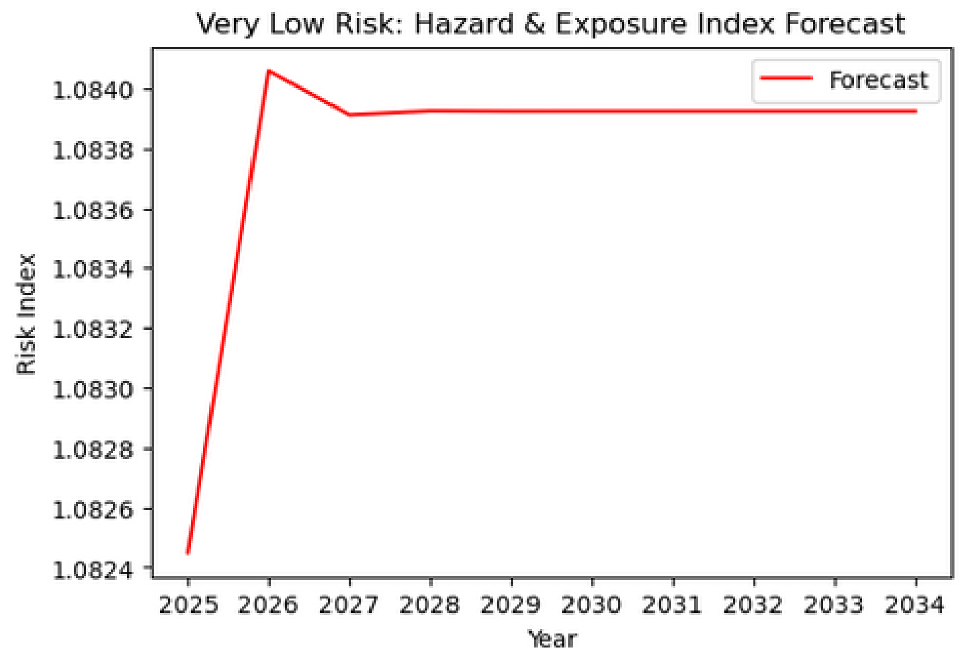
Predict & forecast how INFORM risk data changes over time

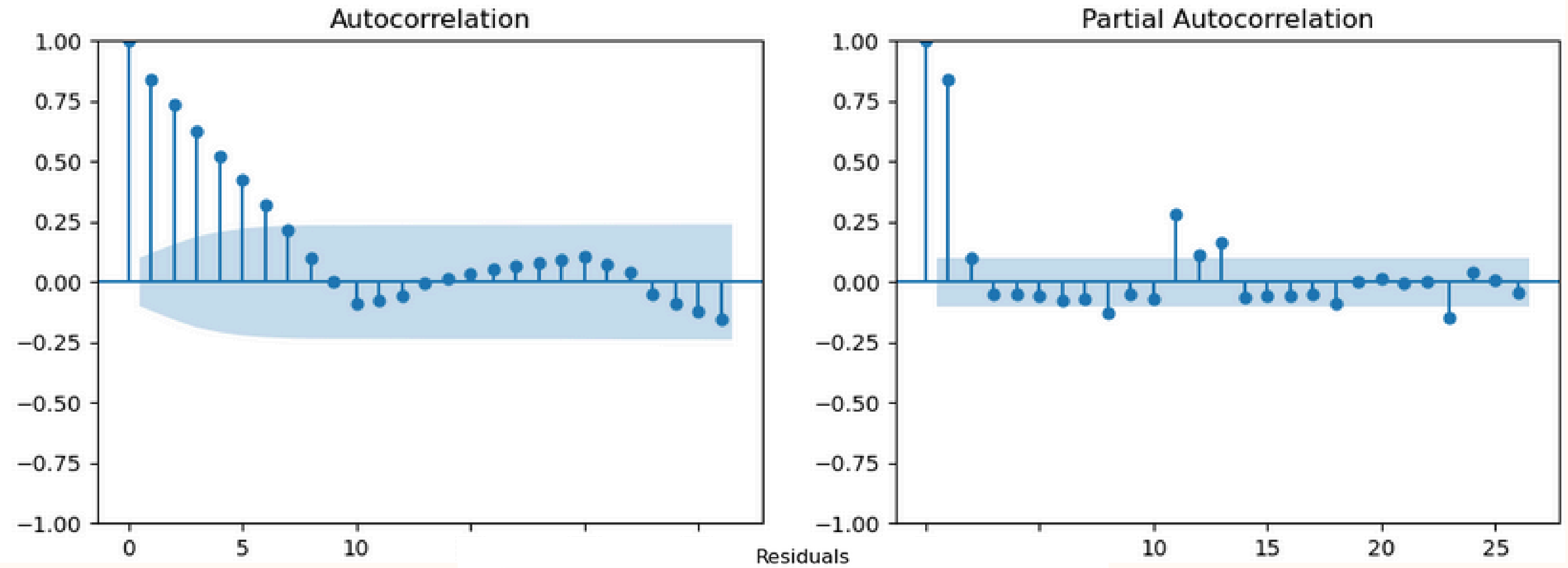
ARIMA Model

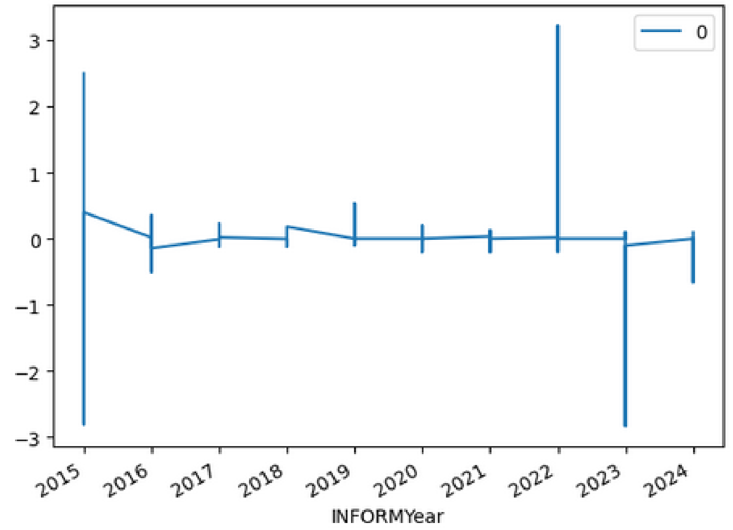
ARIMA Model to forecast Very Low Risk Countries' Hazard & Exposure Risk indices:

Back to Agenda







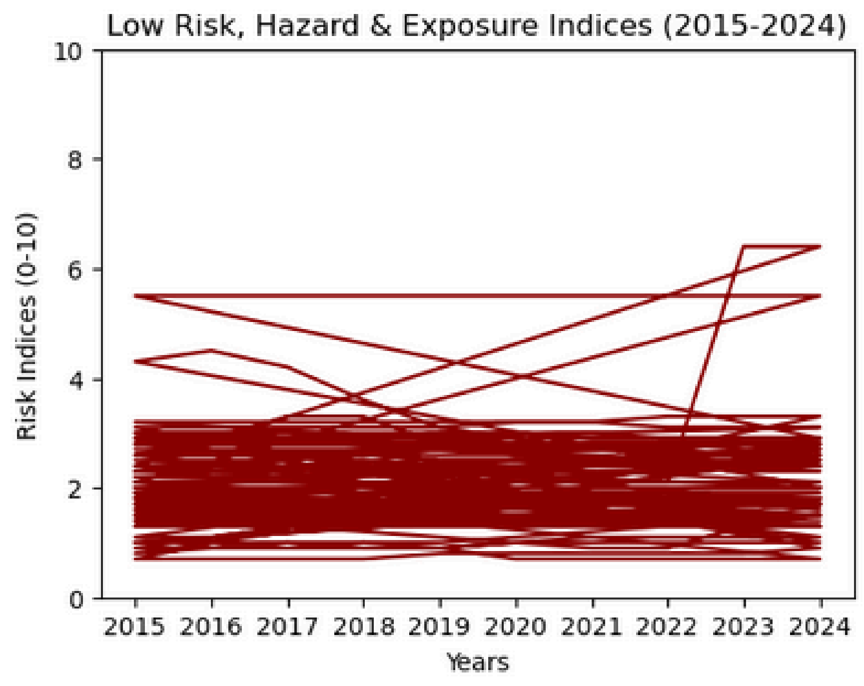


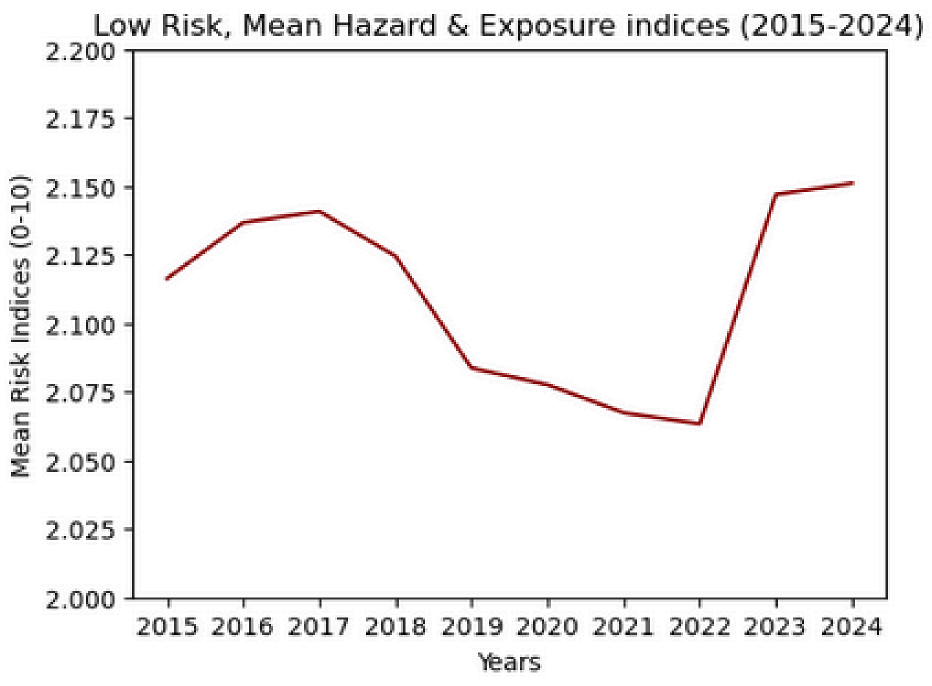

```
Mean Squared Error (MSE): 0.007628361107492l4
Mean Absolute Error (MAE): 0.08343438748863088
Root Mean Squared Error (RMSE): 0.08734048950797185
```

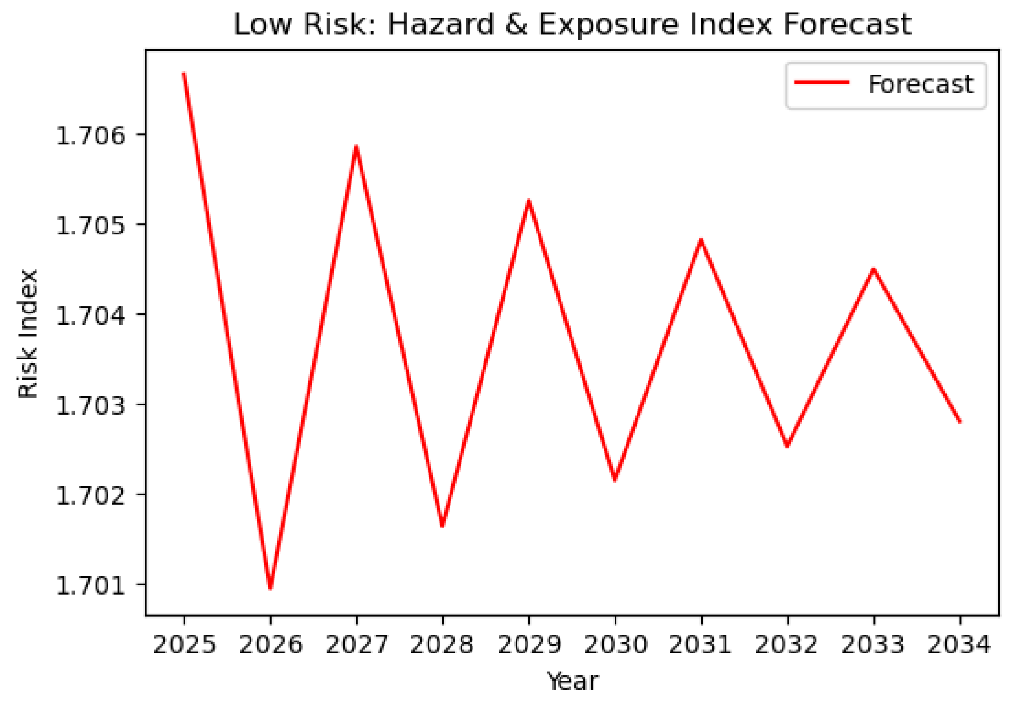
```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Hazard & Exposure Index   No. Observations:            380
Model:                      ARIMA(1, 1, 1)   Log Likelihood          -168.348
Date:                     Mon, 08 Apr 2024   AIC                      342.697
Time:                             03:50:22   BIC                      354.509
Sample:                                  0   HQIC                     347.384
                                     - 380
Covariance Type:                       opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.0916      0.314     -0.292      0.770      -0.706       0.523
ma.L1         -0.0865      0.315     -0.275      0.783      -0.704       0.531
sigma2         0.1423      0.003     51.268      0.000       0.137       0.148
==============================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):          16801.84
Prob(Q):                              0.99   Prob(JB):                      0.00
Heteroskedasticity (H):               0.54   Skew:                         -0.21
Prob(H) (two-sided):                  0.00   Kurtosis:                     35.62
==============================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

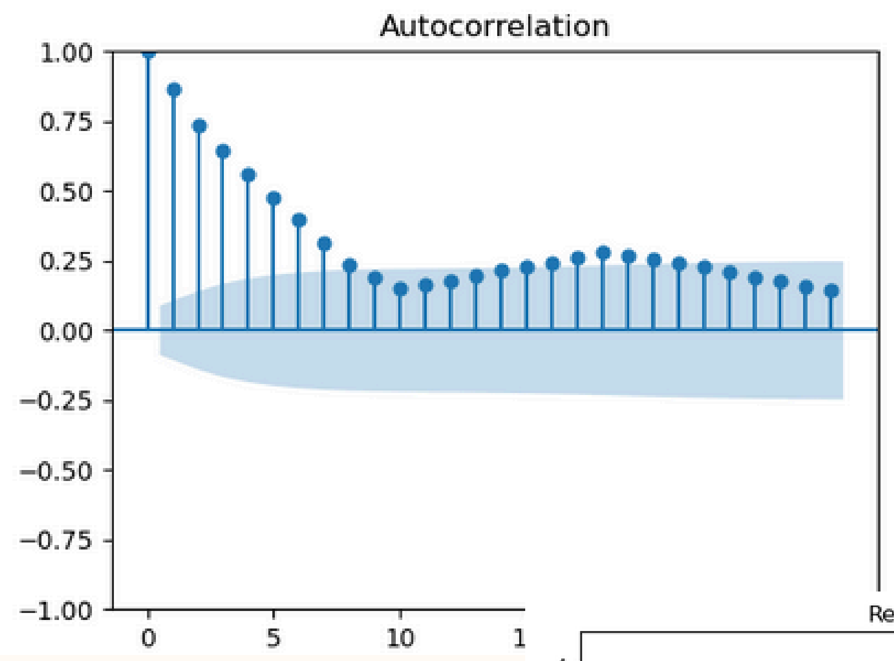
ARIMA Model to forecast Low Risk Countries' Hazard & Exposure Risk indices:

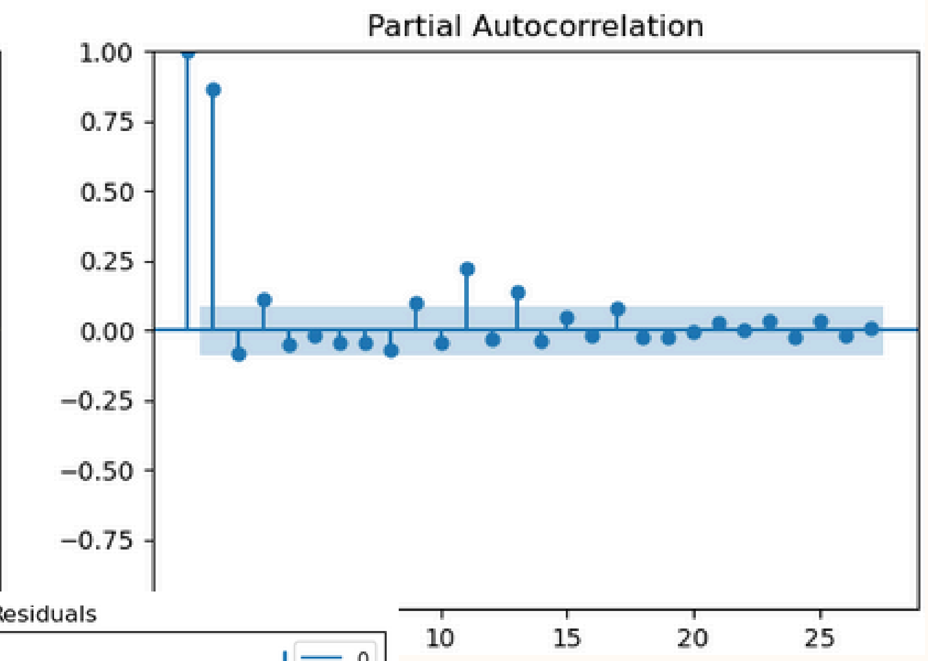
Back to Agenda

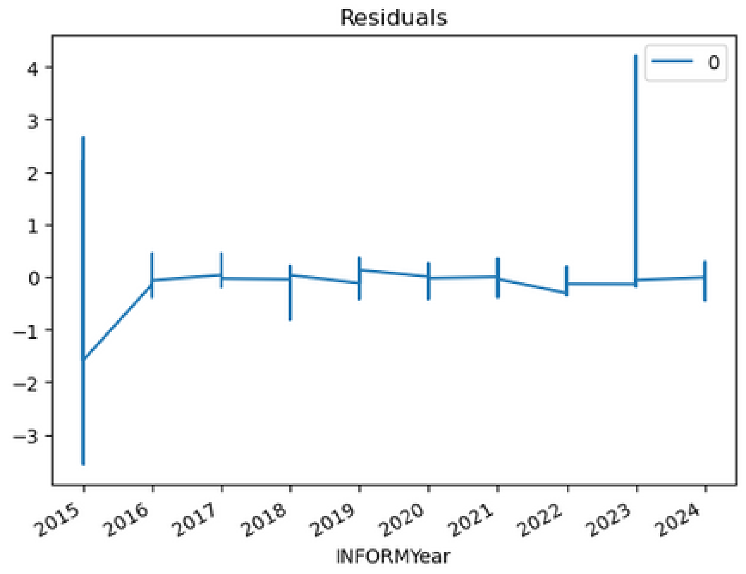










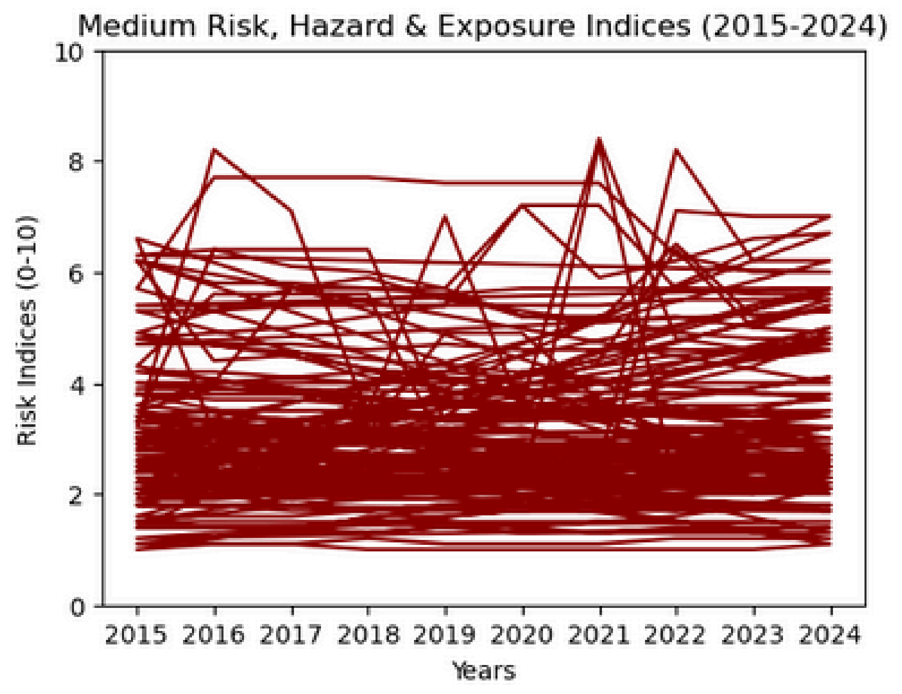


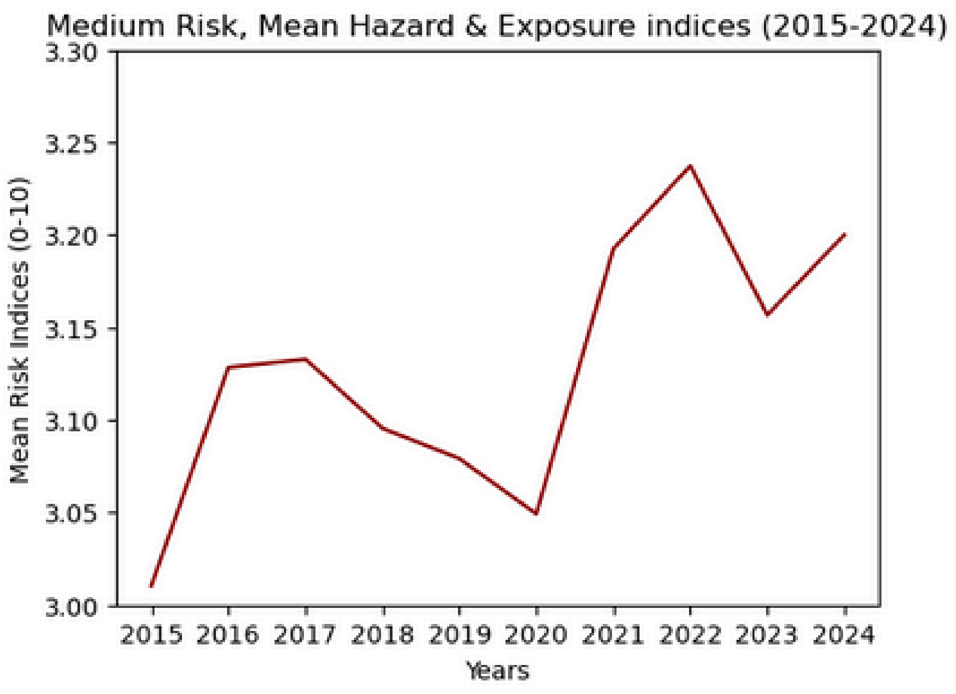

```
Mean Squared Error (MSE): 0.0054238845169744315
Mean Absolute Error (MAE): 0.05371051640434208
Root Mean Squared Error (RMSE): 0.07364702653179171
```

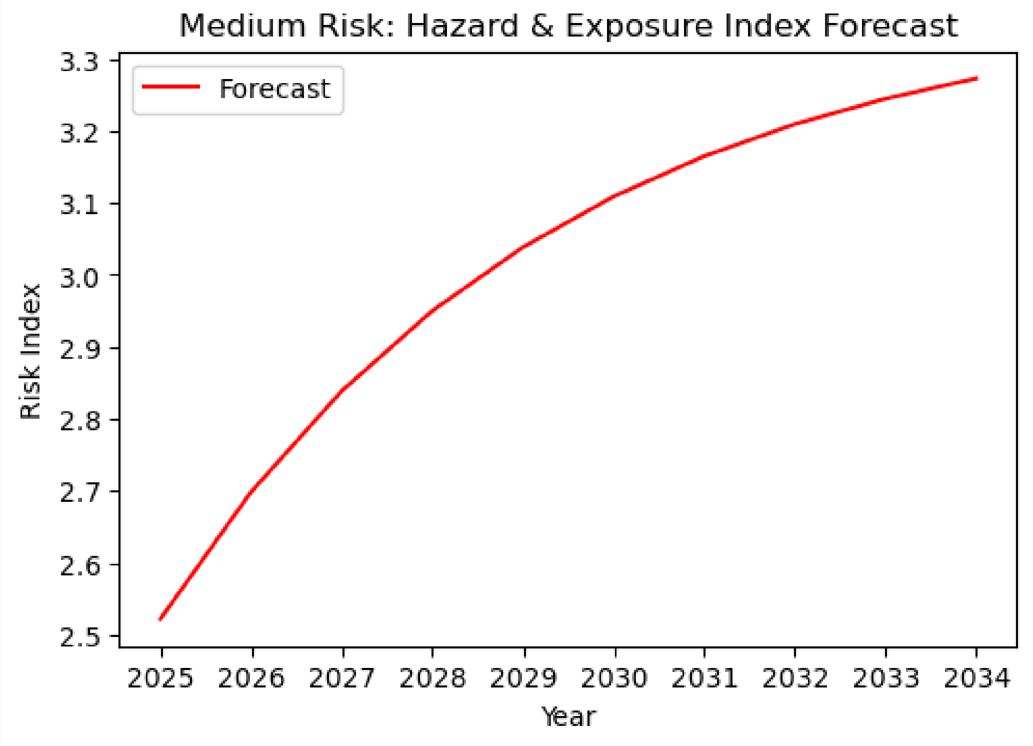
```
                                SARIMAX Results
==============================================================================
Dep. Variable:     Hazard & Exposure Index   No. Observations:            490
Model:                      ARIMA(1, 1, 1)   Log Likelihood          -291.292
Date:                     Mon, 08 Apr 2024   AIC                      588.583
Time:                             03:54:48   BIC                      601.160
Sample:                                  0   HQIC                     593.523
                                     - 490
Covariance Type:                       opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.8590      0.034    -25.280      0.000      -0.926      -0.792
ma.L1          0.9842      0.010     96.079      0.000       0.964       1.004
sigma2         0.1923      0.005     38.131      0.000       0.182       0.202
==============================================================================
Ljung-Box (L1) (Q):                   1.43   Jarque-Bera (JB):          29068.32
Prob(Q):                              0.23   Prob(JB):                      0.00
Heteroskedasticity (H):               1.36   Skew:                          0.54
Prob(H) (two-sided):                  0.05   Kurtosis:                     40.76
==============================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step)
```

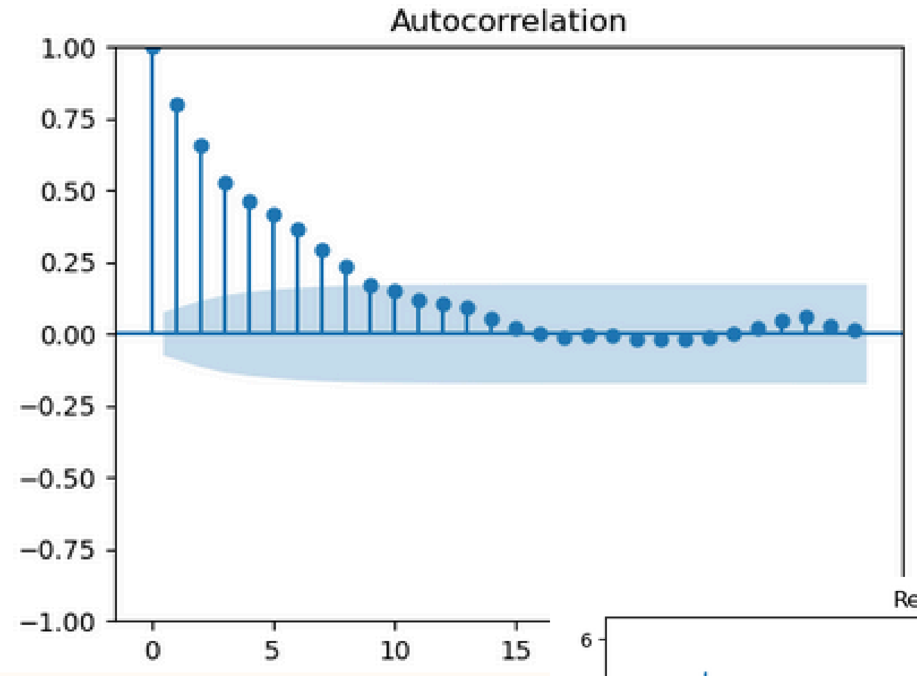
ARIMA Model to forecast Medium Risk Countries' Hazard & Exposure Risk indices:

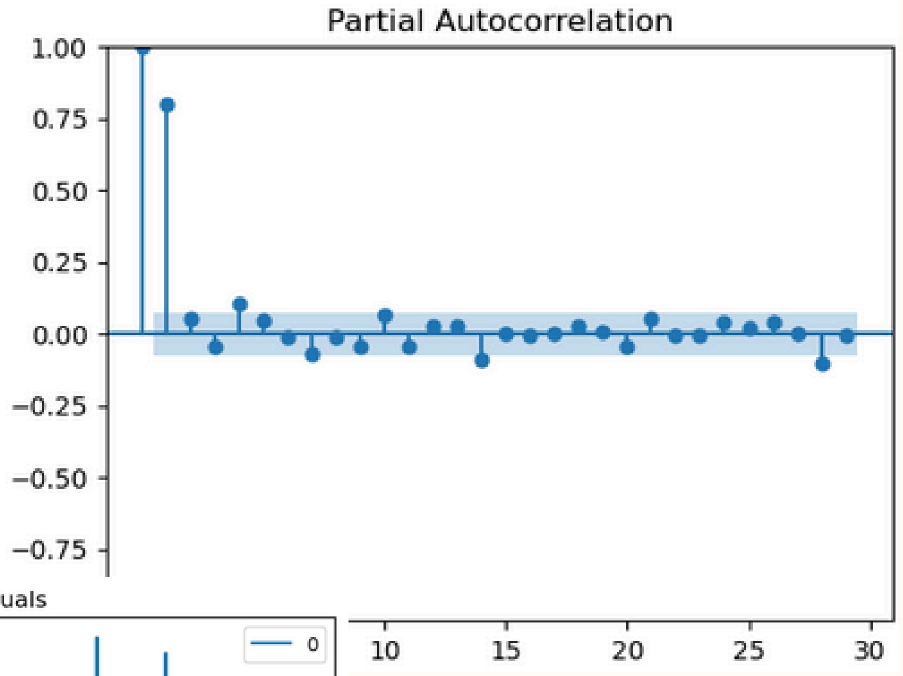
Back to Agenda









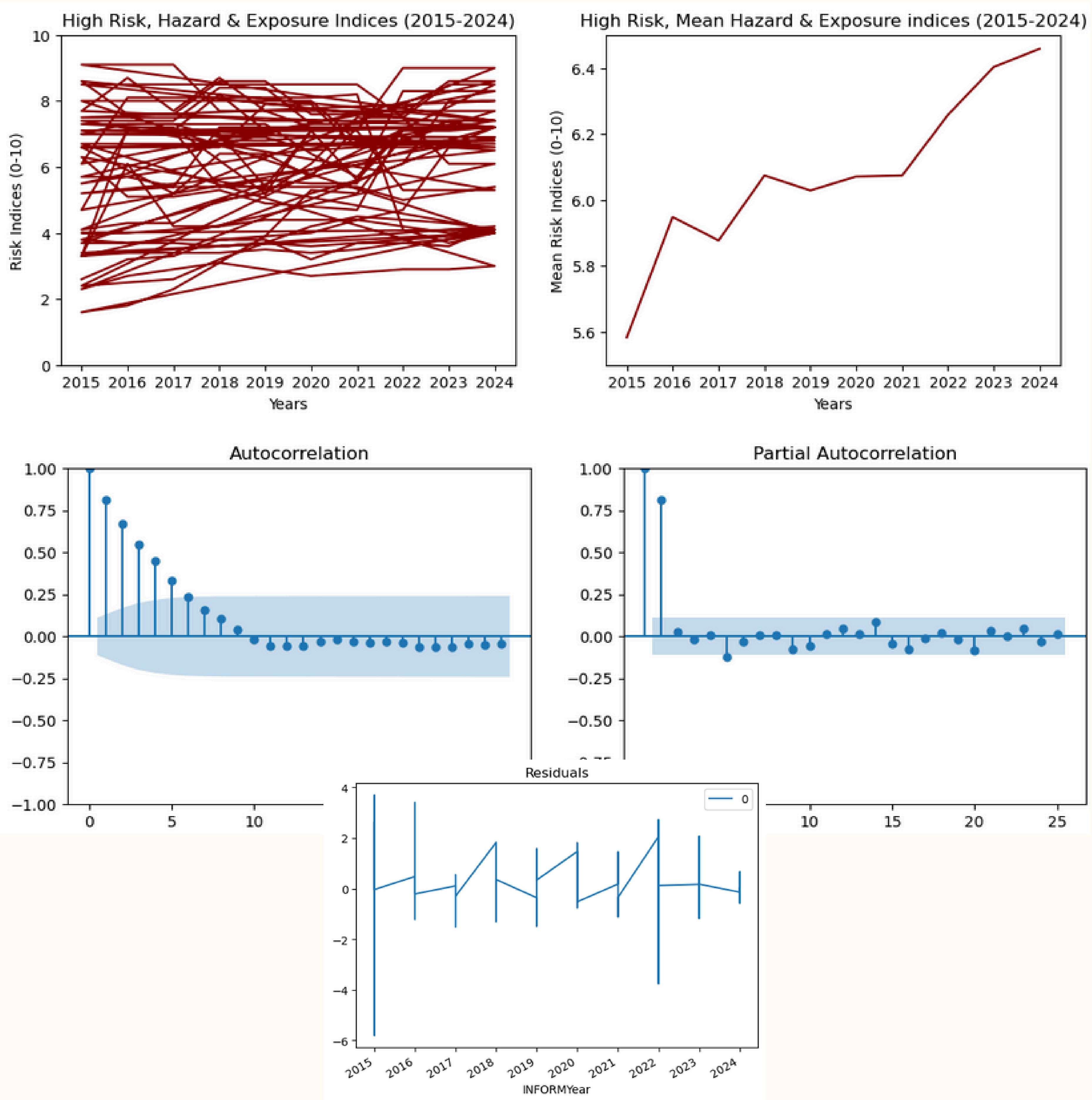


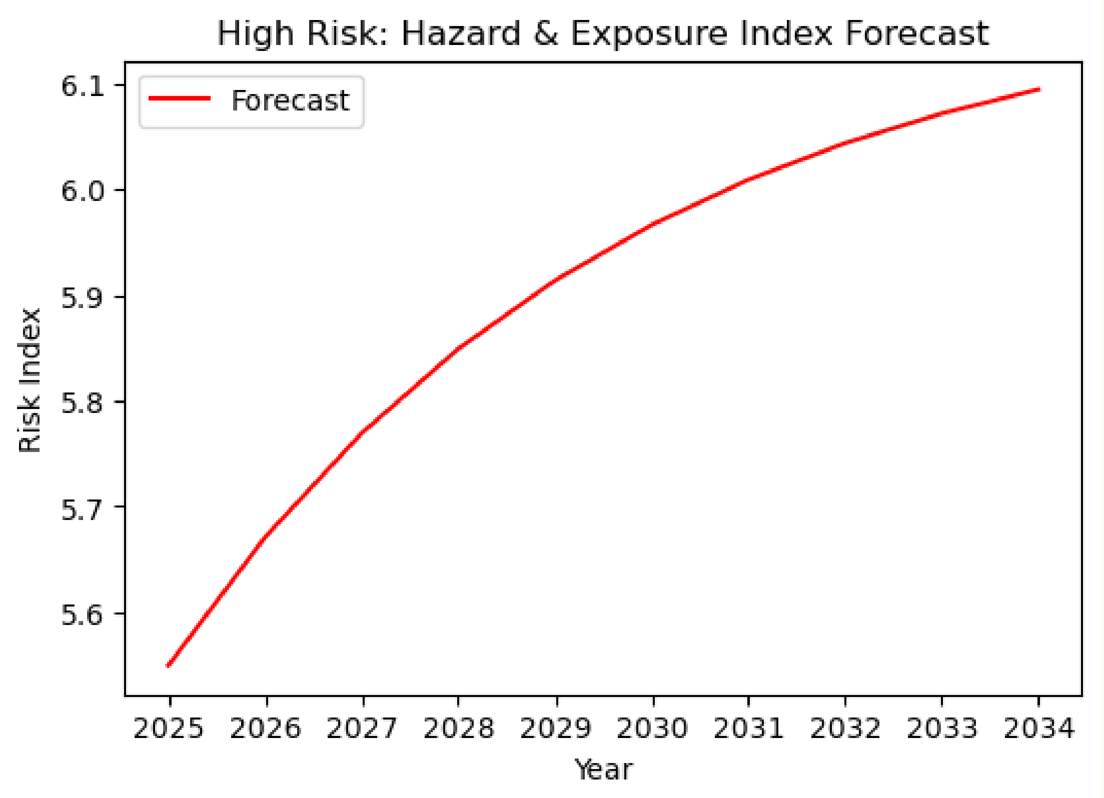

```
Mean Squared Error (MSE): 0.5499464852573603
Mean Absolute Error (MAE): 0.7147981896867188
Root Mean Squared Error (RMSE): 0.7415837681997633
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Hazard & Exposure Index   No. Observations:          670
Model:                      ARIMA(1, 1, 1)   Log Likelihood        -869.607
Date:                     Mon, 08 Apr 2024   AIC                   1745.213
Time:                             04:07:43   BIC                   1758.731
Sample:                                  0   HQIC                  1750.449
                                     - 670
Covariance Type:                       opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.7950      0.021     38.299      0.000       0.754       0.836
ma.L1         -0.9936      0.007   -144.158      0.000      -1.007      -0.980
sigma2         0.7854      0.018     43.496      0.000       0.750       0.821
==============================================================================
Ljung-Box (L1) (Q):                   1.15   Jarque-Bera (JB):          6510.32
Prob(Q):                              0.28   Prob(JB):                     0.00
Heteroskedasticity (H):               1.09   Skew:                         1.04
Prob(H) (two-sided):                  0.51   Kurtosis:                    18.14
==============================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

ARIMA Model to forecast High Risk Countries' Hazard & Exposure Risk indices:

Back to Agenda



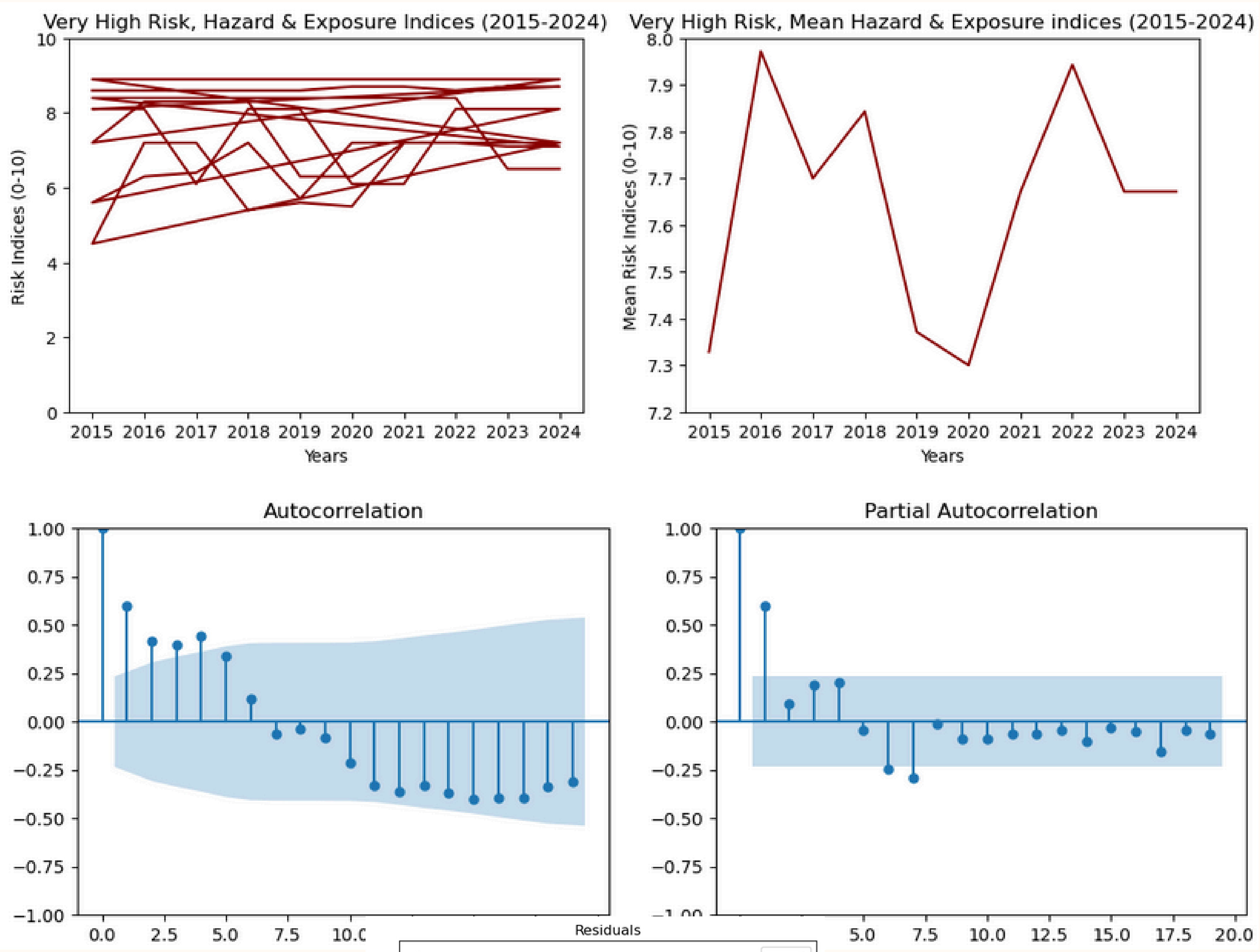


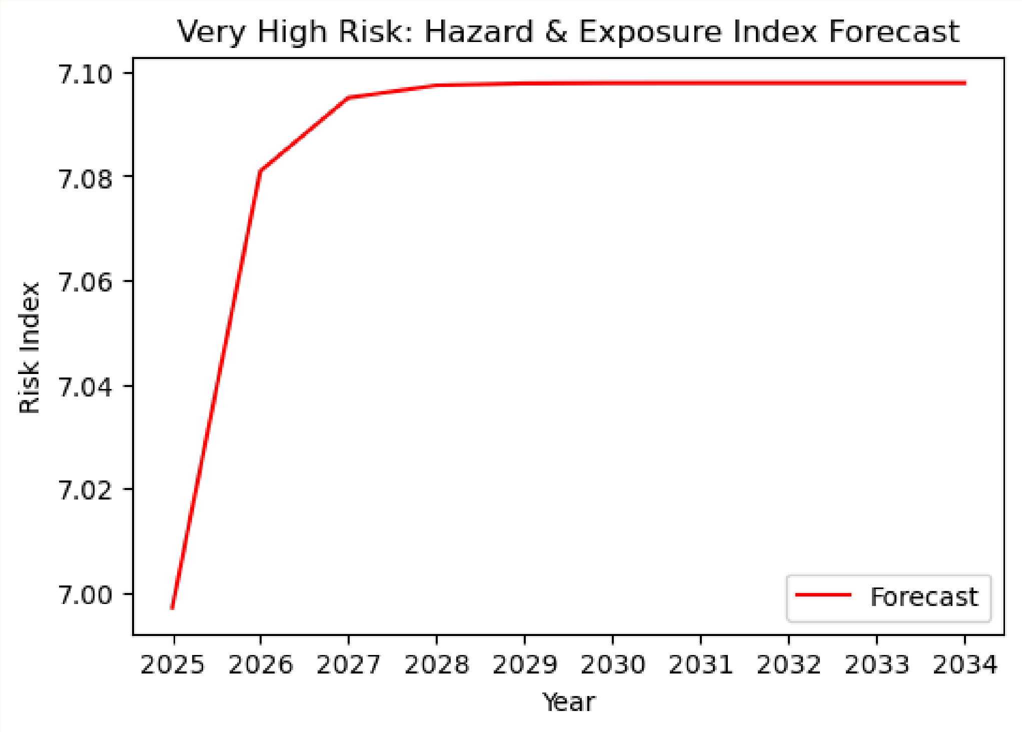

```
Mean Squared Error (MSE): 0.9997580409036935
Mean Absolute Error (MAE): 0.9460648697163029
Root Mean Squared Error (RMSE): 0.9998790131329357
```

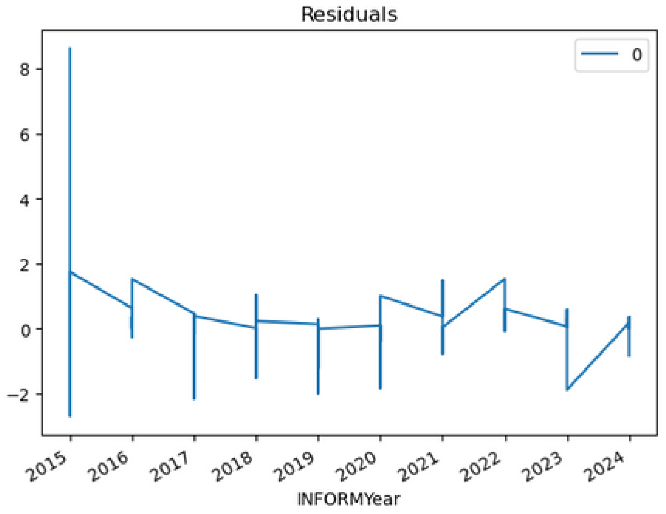
```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Hazard & Exposure Index   No. Observations:            310
Model:                      ARIMA(1, 1, 1)   Log Likelihood           -445.771
Date:                     Mon, 08 Apr 2024   AIC                       897.543
Time:                             04:04:27   BIC                       908.743
Sample:                                  0   HQIC                      902.021
                                     - 310
Covariance Type:                       opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8114      0.057     14.253      0.000       0.700       0.923
ma.L1         -0.9861      0.014    -72.257      0.000      -1.013      -0.959
sigma2         1.0436      0.064     16.432      0.000       0.919       1.168
==============================================================================
Ljung-Box (L1) (Q):                 0.28   Jarque-Bera (JB):               844.79
Prob(Q):                            0.60   Prob(JB):                         0.00
Heteroskedasticity (H):             1.19   Skew:                            -1.16
Prob(H) (two-sided):                0.37   Kurtosis:                        10.76
==============================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

ARIMA Model to forecast Very High Risk Countries' Hazard & Exposure Risk indices:

Back to Agenda





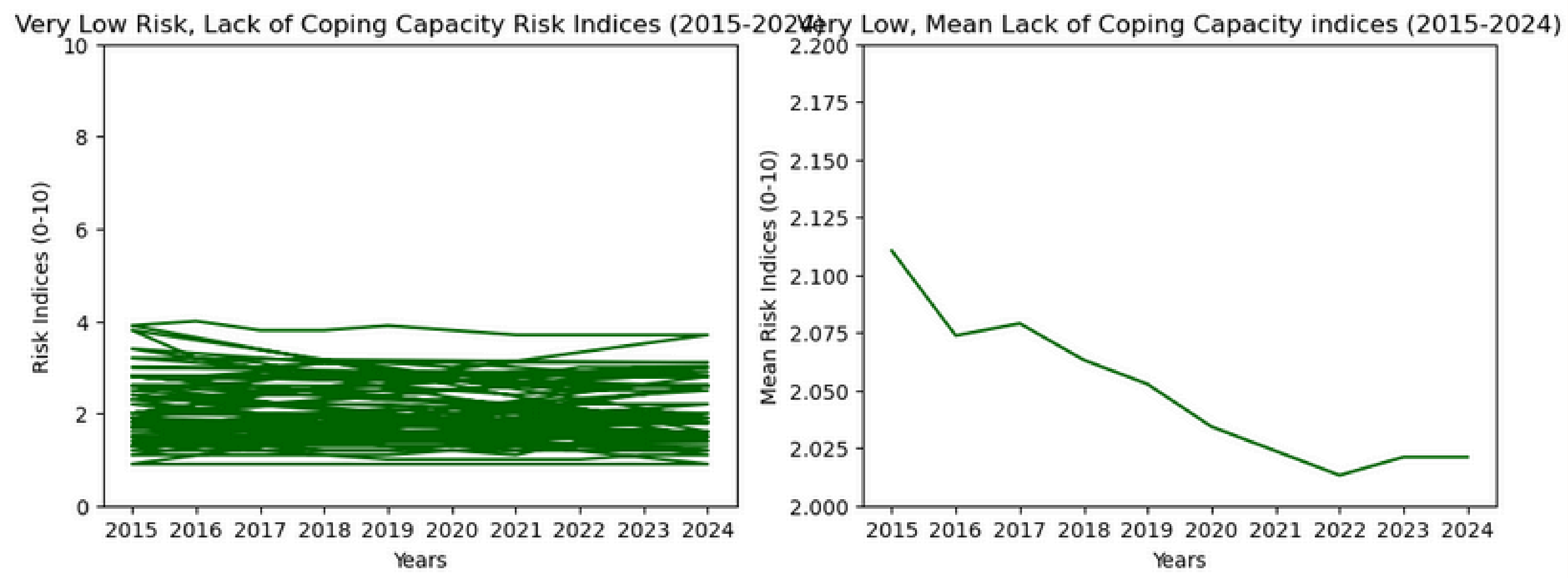


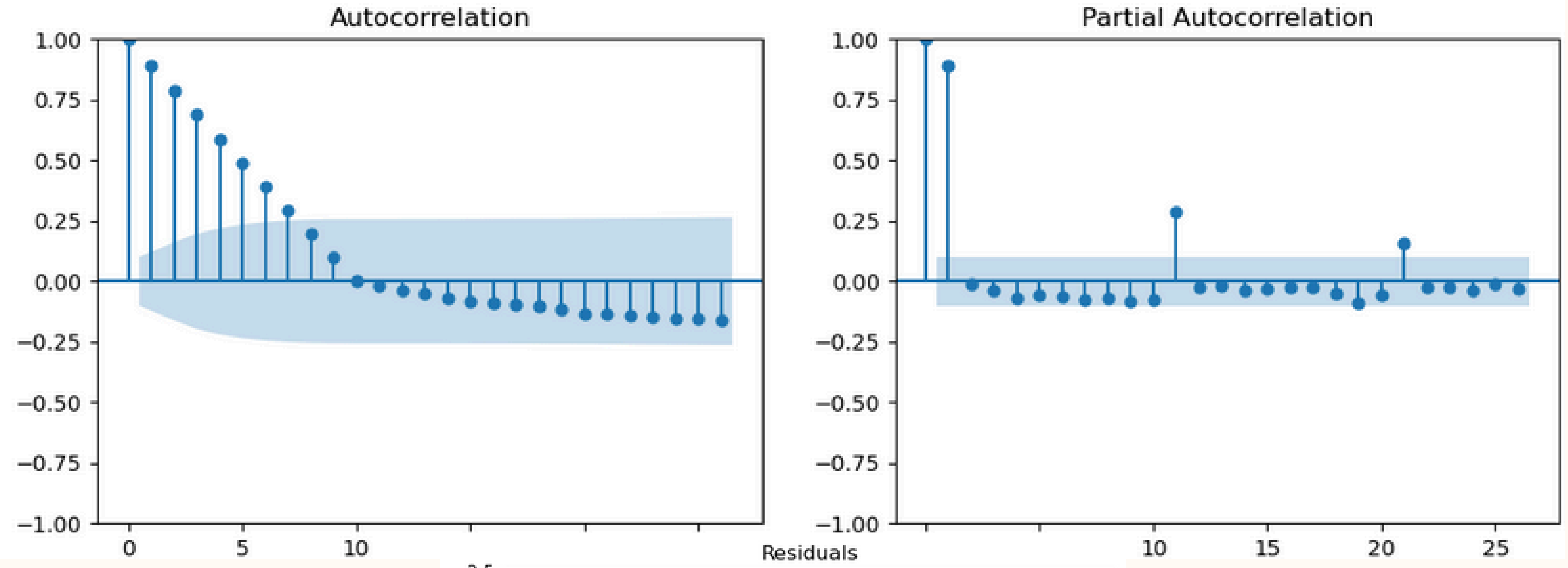

```
Mean Squared Error (MSE): 1.460450256613416
Mean Absolute Error (MAE): 1.1733704281047608
Root Mean Squared Error (RMSE): 1.2084909005091
```

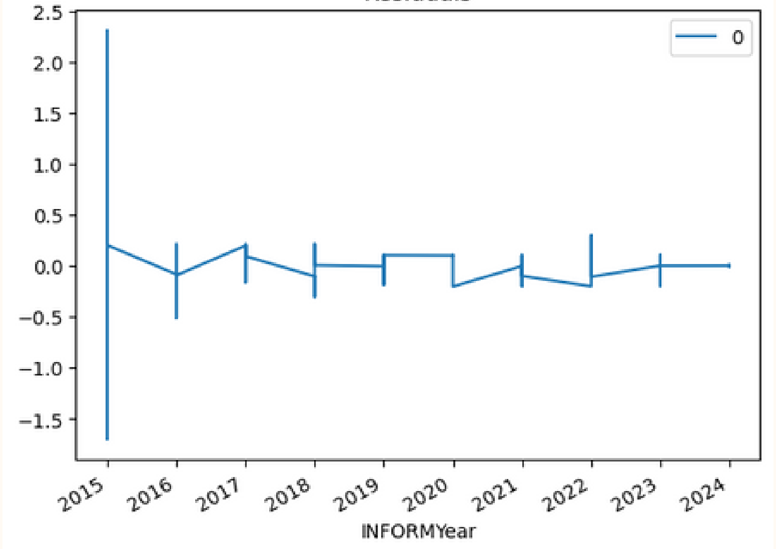
```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Hazard & Exposure Index   No. Observations:               70
Model:                      ARIMA(1, 1, 1)   Log Likelihood             -90.340
Date:                     Mon, 08 Apr 2024   AIC                        186.679
Time:                             04:00:52   BIC                        193.382
Sample:                                  0   HQIC                       189.338
                                      - 70
Covariance Type:                       opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.1684      0.349      0.483      0.629      -0.515       0.852
ma.L1         -0.6062      0.294     -2.060      0.039      -1.183      -0.029
sigma2         0.7999      0.111      7.204      0.000       0.582       1.018
==============================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                12.62
Prob(Q):                              0.92   Prob(JB):                         0.00
Heteroskedasticity (H):               1.17   Skew:                            -0.89
Prob(H) (two-sided):                  0.71   Kurtosis:                         4.12
==============================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

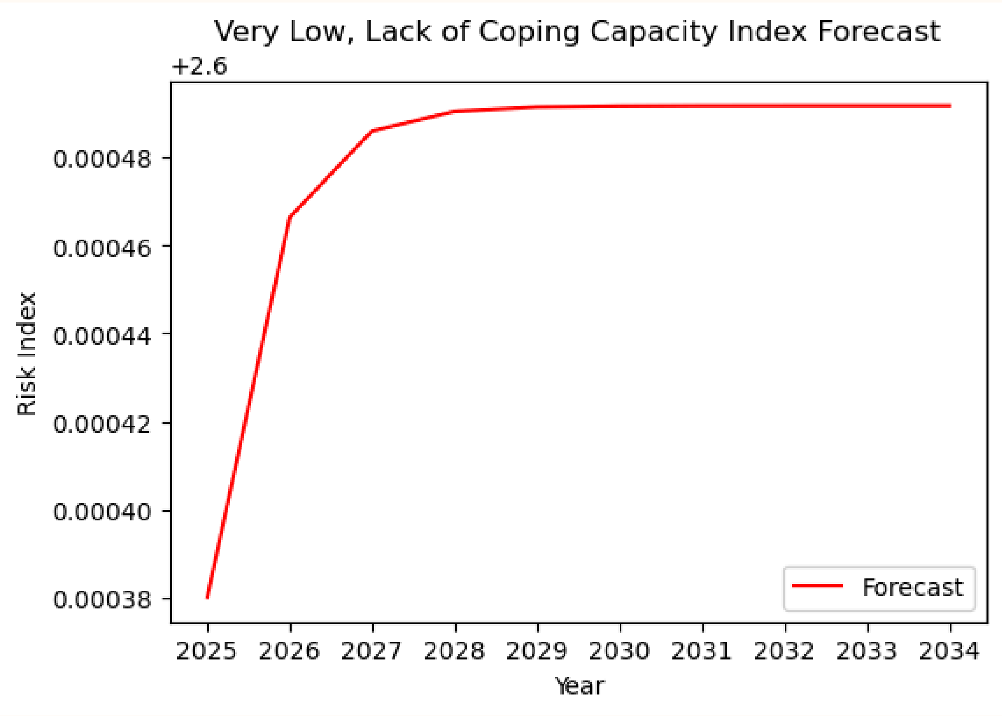
ARIMA Model to forecast Very Low Risk Countries' Lack of Coping Capacity Risk indices:

Back to Agenda







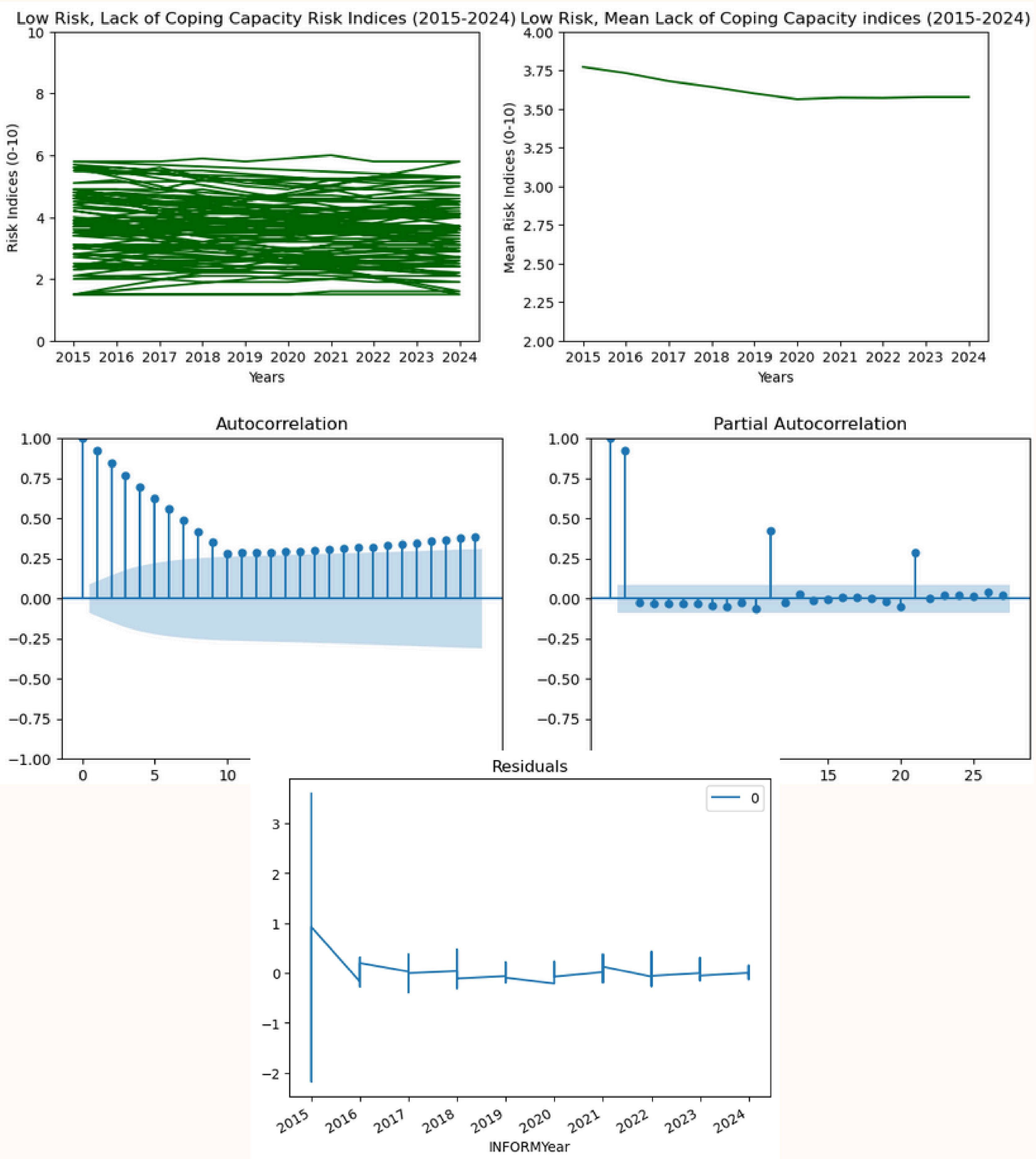


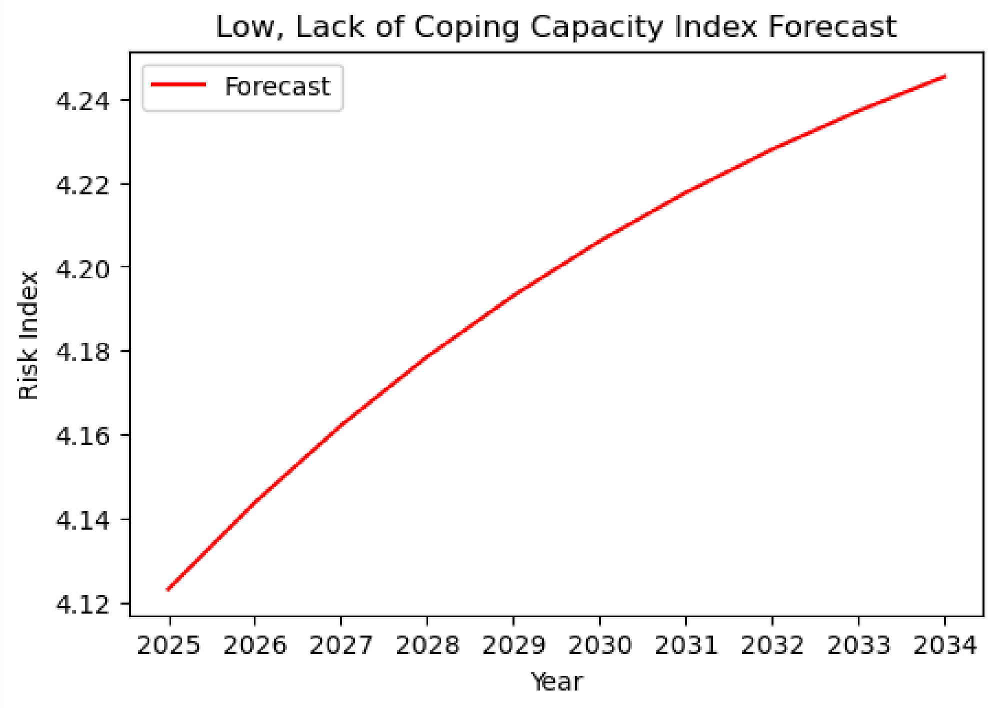

```
Mean Squared Error (MSE): 0.021887986255368 35
Mean Absolute Error (MAE): 0.11981774778535566
Root Mean Squared Error (RMSE): 0.14794588961971317
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Lack of Coping Capacity Index   No. Observations:            380
Model:                            ARIMA(1, 1, 1)   Log Likelihood          -109.645
Date:                           Mon, 08 Apr 2024   AIC                      225.289
Time:                                   04:12:57   BIC                      237.102
Sample:                                        0   HQIC                     229.977
                                           - 380
Covariance Type:                             opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.2271      2.833      0.080      0.936      -5.326       5.780
ma.L1         -0.2742      2.823     -0.097      0.923      -5.807       5.258
sigma2         0.1044      0.002     46.574      0.000       0.100       0.109
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              6981.63
Prob(Q):                              0.99   Prob(JB):                         0.00
Heteroskedasticity (H):               1.87   Skew:                             1.16
Prob(H) (two-sided):                  0.00   Kurtosis:                        23.90
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

ARIMA Model to forecast Low Risk Countries' Lack of Coping Capacity Risk indices:

Back to Agenda



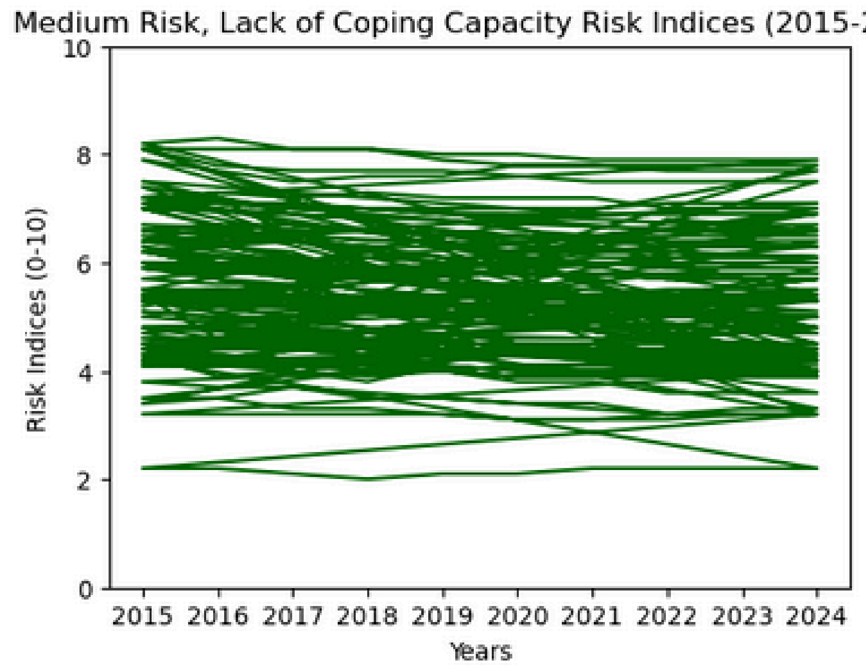


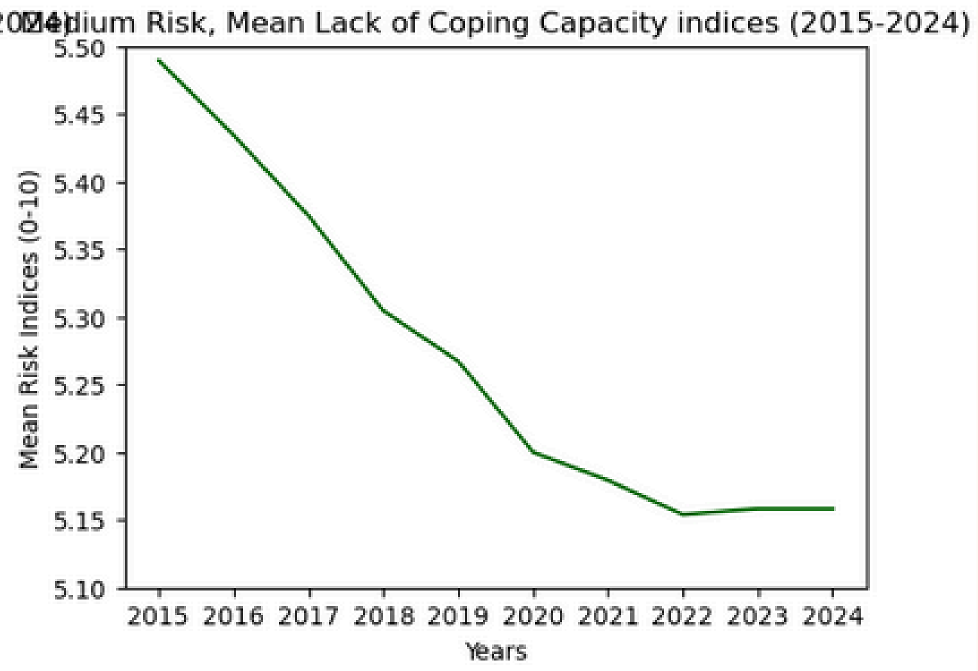

```
Mean Squared Error (MSE): 0.018059835212866686
Mean Absolute Error (MAE): 0.12145136003107489
Root Mean Squared Error (RMSE): 0.1343868863128642
```

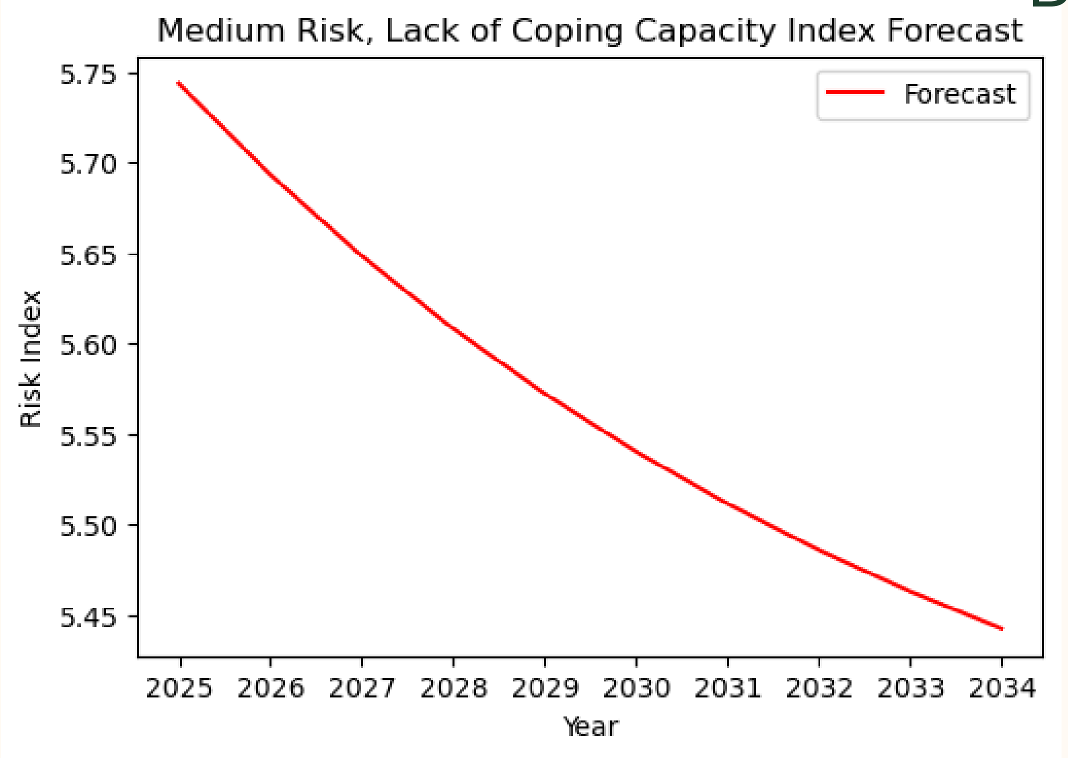
```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Lack of Coping Capacity Index   No. Observations:        490
Model:                            ARIMA(1, 1, 1)   Log Likelihood      -250.870
Date:                           Tue, 09 Apr 2024   AIC                  507.741
Time:                                   00:12:24   BIC                  520.318
Sample:                                        0   HQIC                 512.681
                                           - 490
Covariance Type:                             opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8909      0.044     20.217      0.000       0.805       0.977
ma.L1         -0.9846      0.012    -85.518      0.000      -1.007      -0.962
sigma2         0.1631      0.006     28.778      0.000       0.152       0.174
===================================================================================
Ljung-Box (L1) (Q):                   0.64   Jarque-Bera (JB):              5350.98
Prob(Q):                              0.42   Prob(JB):                         0.00
Heteroskedasticity (H):               1.18   Skew:                             1.19
Prob(H) (two-sided):                  0.28   Kurtosis:                        19.03
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

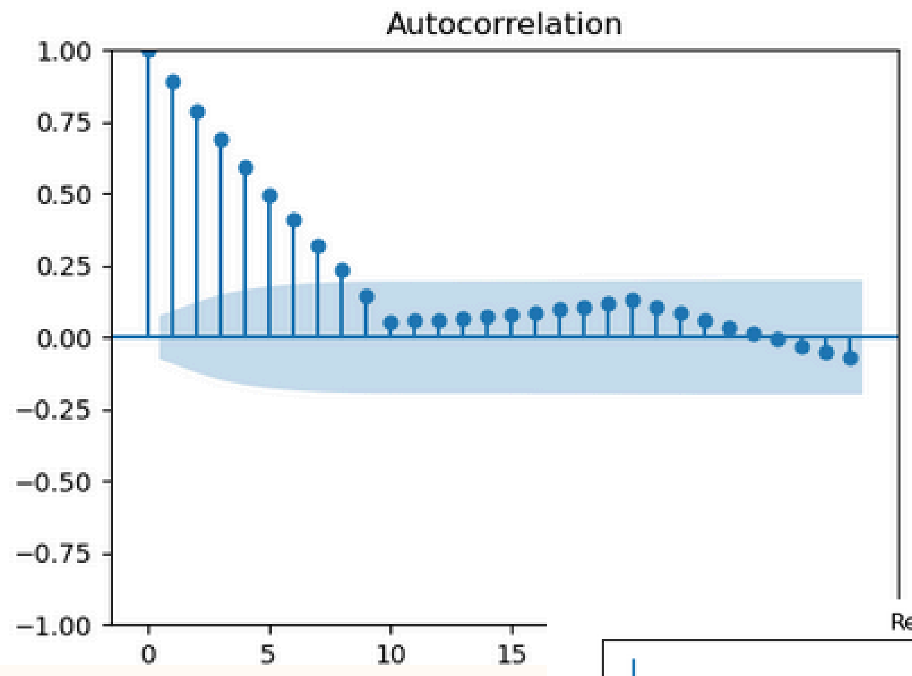
ARIMA Model to forecast Medium Risk Countries' Lack of Coping Capacity Risk indices:

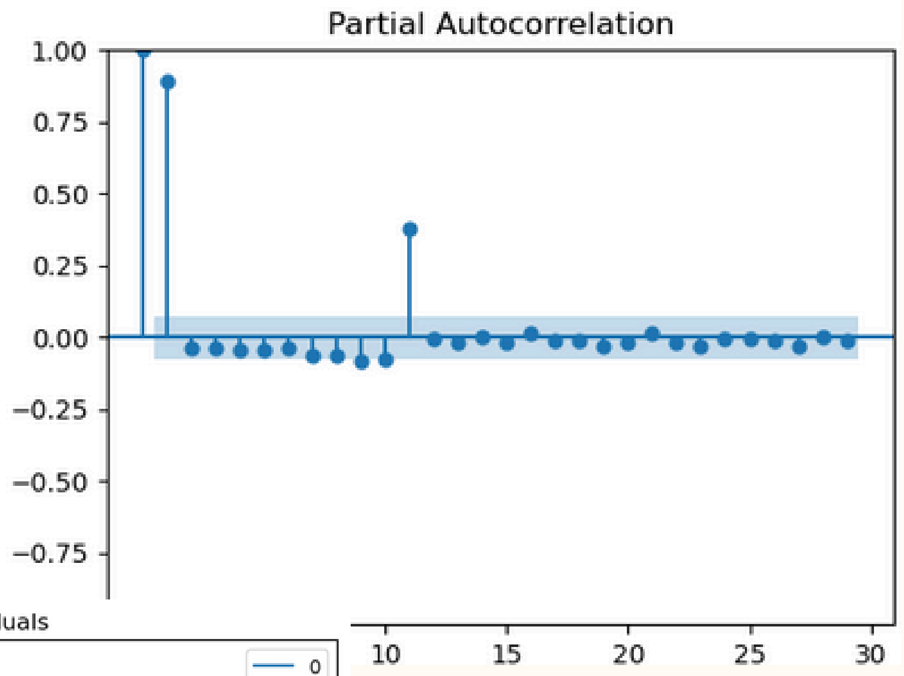
Back to Agenda

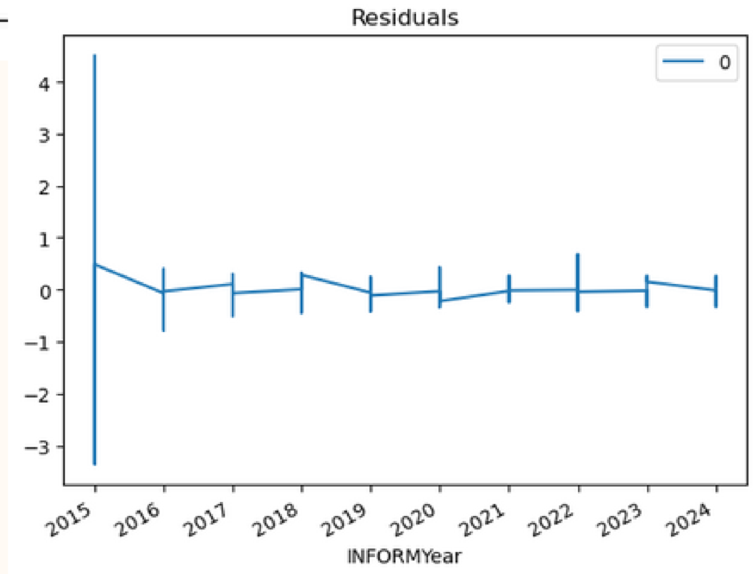










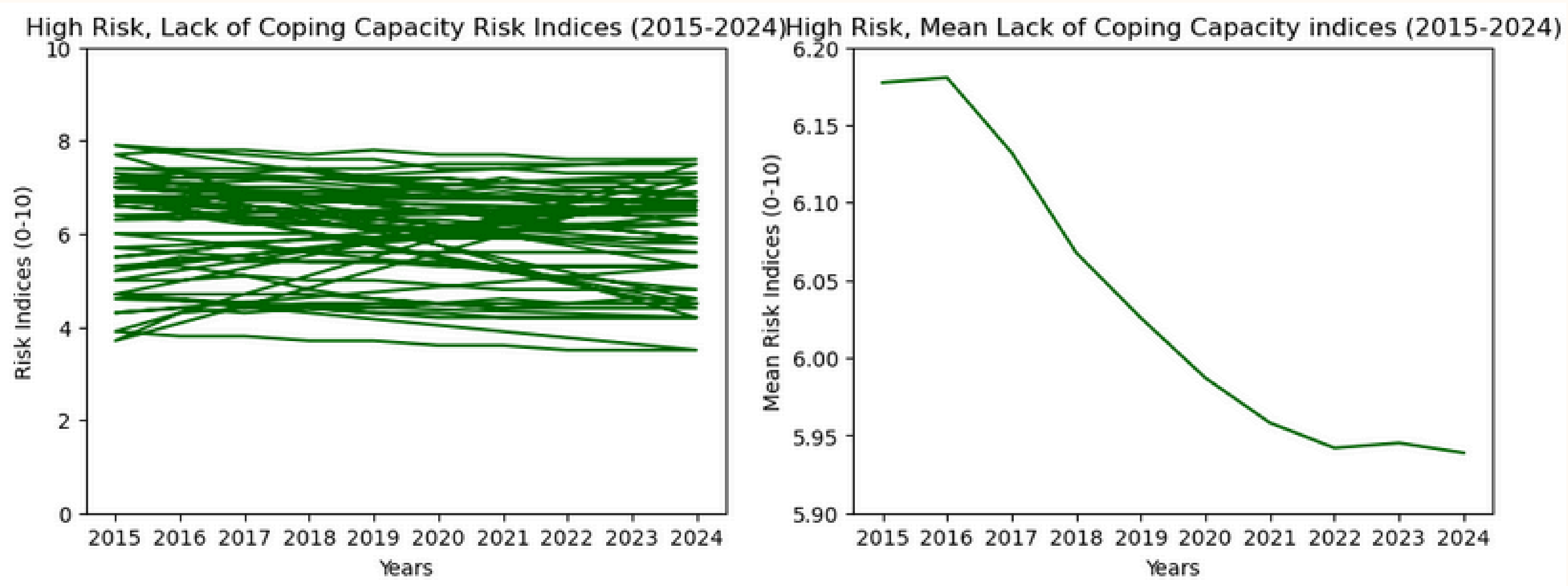


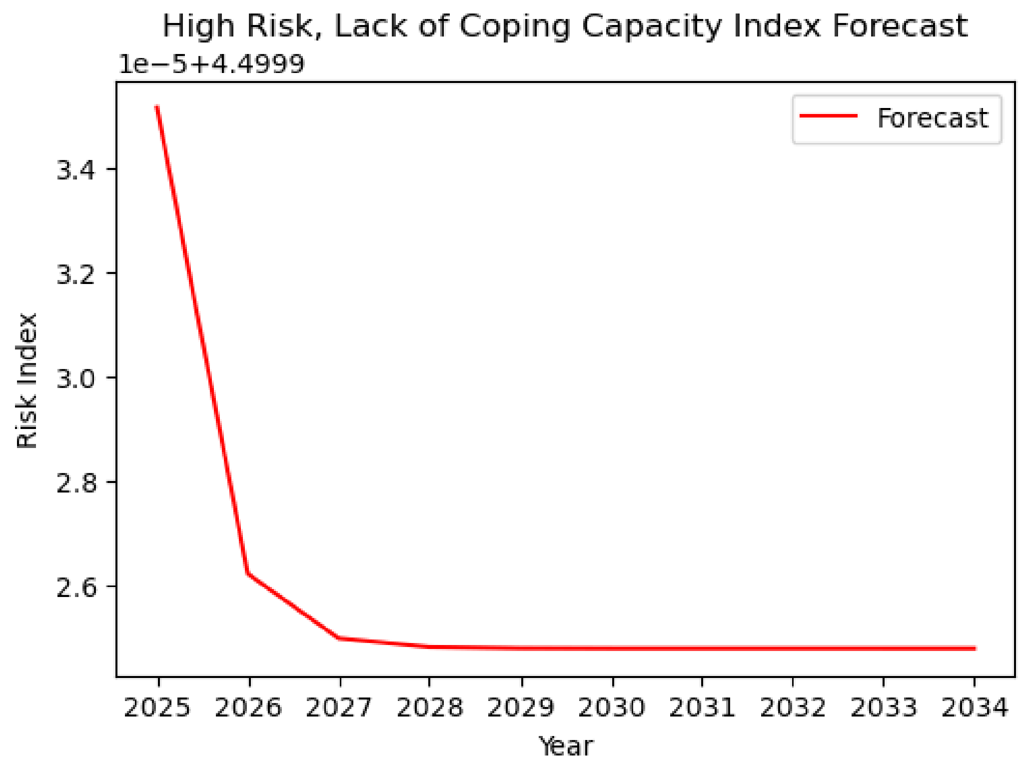

```
Mean Squared Error (MSE): 0.06385586356543467
Mean Absolute Error (MAE): 0.23913521617474282
Root Mean Squared Error (RMSE): 0.252697177596891
```

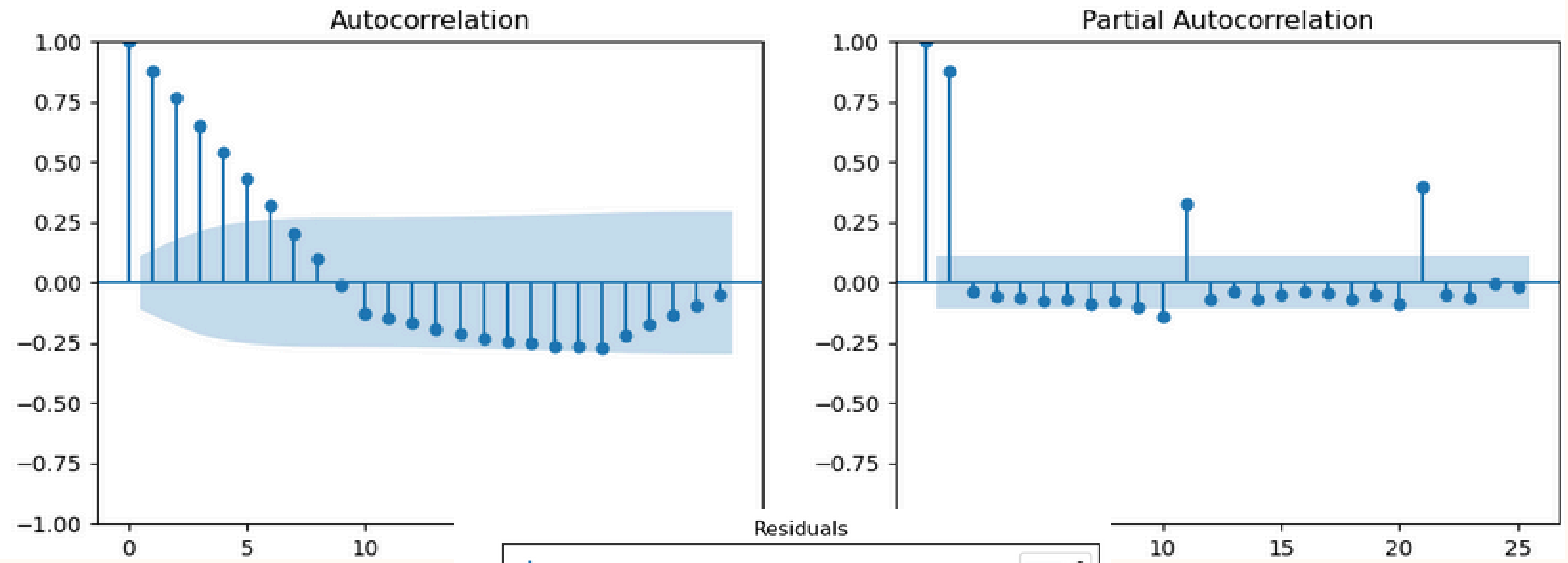
```
                                  SARIMAX Results
==============================================================================================
Dep. Variable:     Lack of Coping Capacity Index   No. Observations:                  670
Model:                            ARIMA(1, 1, 1)   Log Likelihood                -579.502
Date:                           Mon, 08 Apr 2024   AIC                           1165.005
Time:                                   04:25:06   BIC                           1178.522
Sample:                                        0   HQIC                          1170.241
                                           - 670
Covariance Type:                             opg
=====================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------
ar.L1          0.8941      0.031     29.019      0.000       0.834       0.954
ma.L1         -1.0000      0.462     -2.167      0.030      -1.905      -0.095
sigma2         0.3293      0.149      2.212      0.027       0.038       0.621
============================================================================================
Ljung-Box (L1) (Q):                   0.68   Jarque-Bera (JB):              11232.73
Prob(Q):                              0.41   Prob(JB):                          0.00
Heteroskedasticity (H):               0.96   Skew:                              2.04
Prob(H) (two-sided):                  0.74   Kurtosis:                         22.66
============================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

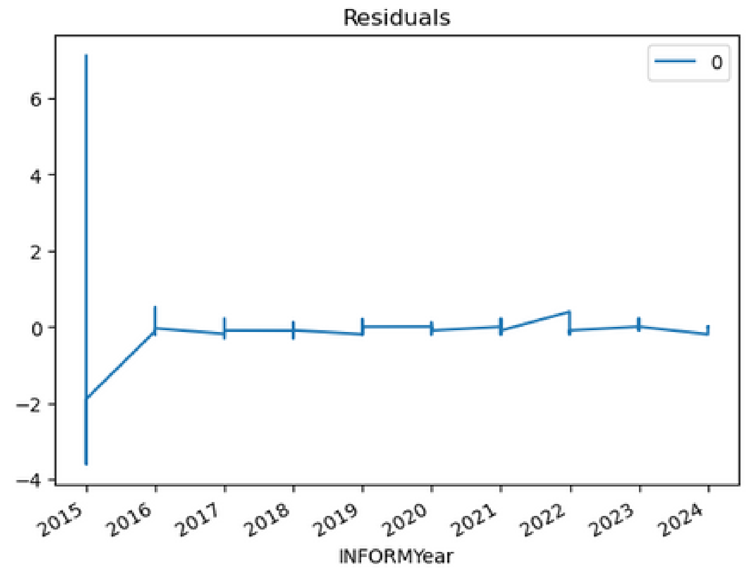
ARIMA Model to forecast High Risk Countries' Lack of Coping Capacity Risk indices:

Back to Agenda







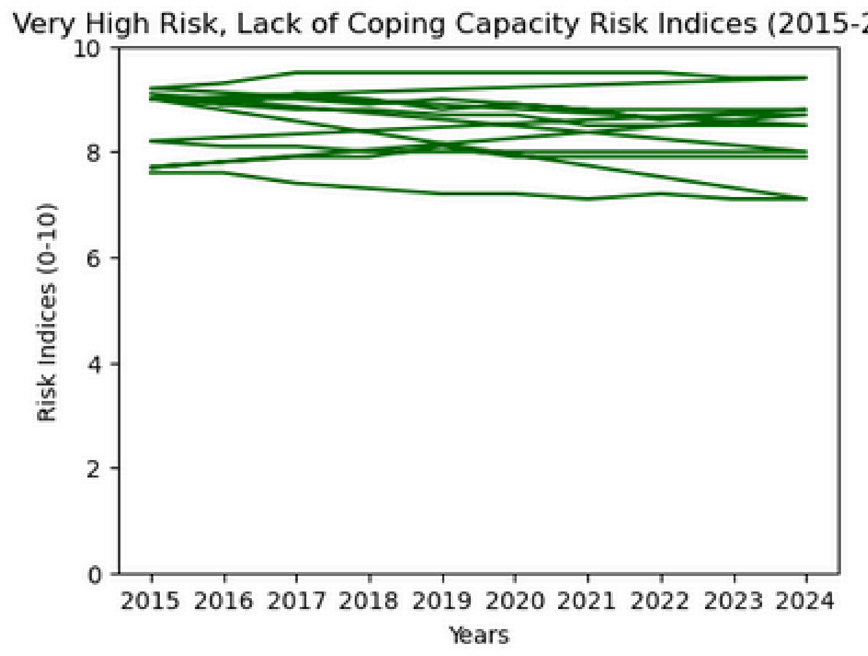


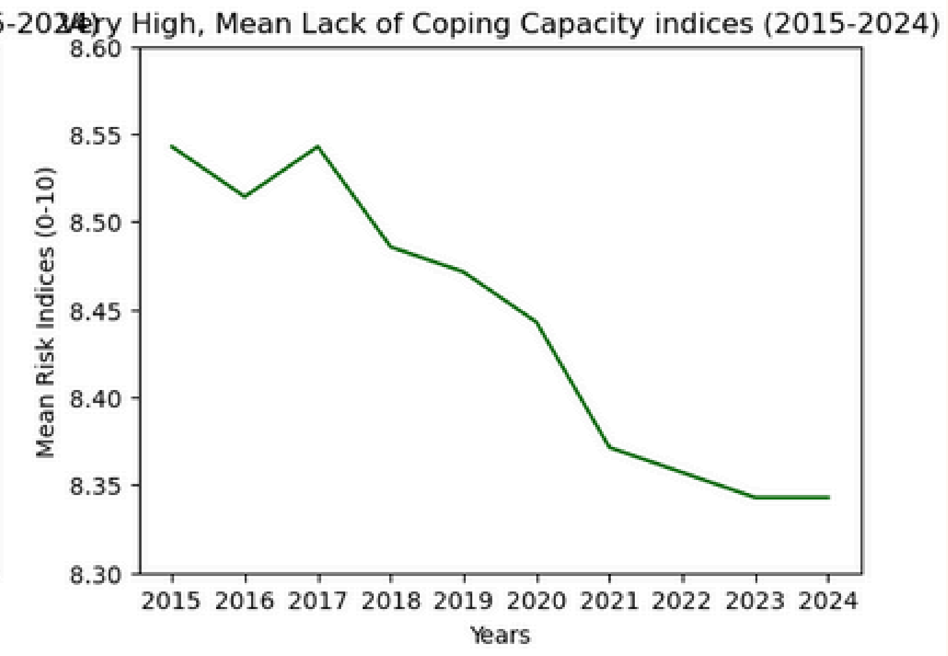

```
Mean Squared Error (MSE): 0.09702738443573523
Mean Absolute Error (MAE): 0.21005897588719077
Root Mean Squared Error (RMSE): 0.31149219000760714
```

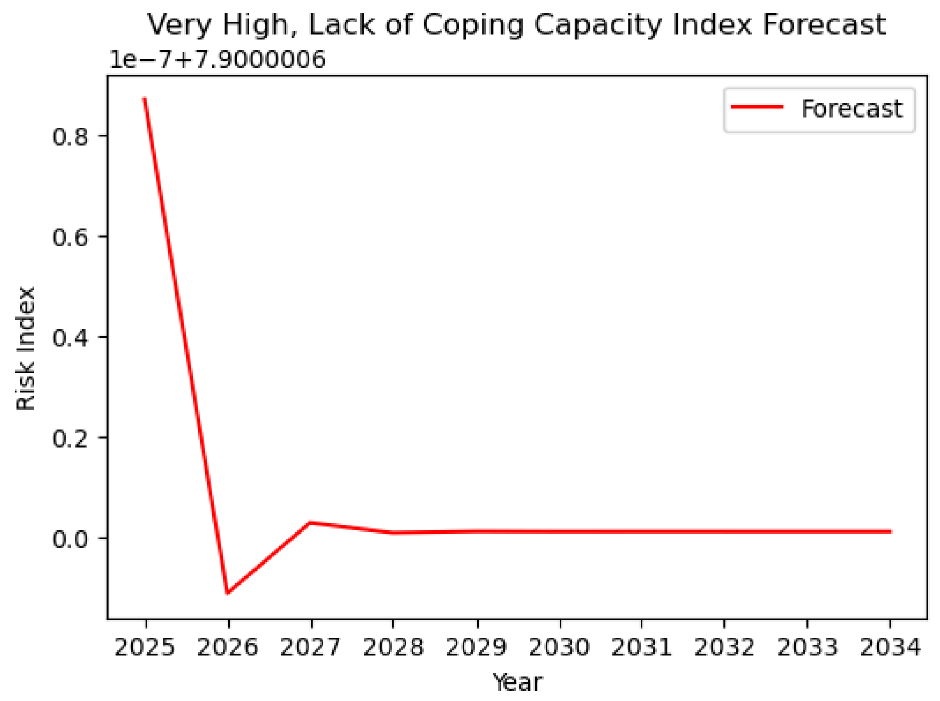
```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Lack of Coping Capacity Index   No. Observations:      310
Model:                        ARIMA(1, 1, 1)   Log Likelihood         -247.462
Date:                       Mon, 08 Apr 2024   AIC                     500.924
Time:                               04:28:24   BIC                     512.124
Sample:                                    0   HQIC                    505.401
                                       - 310
Covariance Type:                         opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.1380      8.637      0.016      0.987     -16.790     17.066
ma.L1         -0.1672      8.626     -0.019      0.985     -17.074     16.740
sigma2         0.2905      0.007     43.221      0.000       0.277      0.304
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              6341.06
Prob(Q):                              1.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.79   Skew:                            -0.10
Prob(H) (two-sided):                  0.24   Kurtosis:                        25.19
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

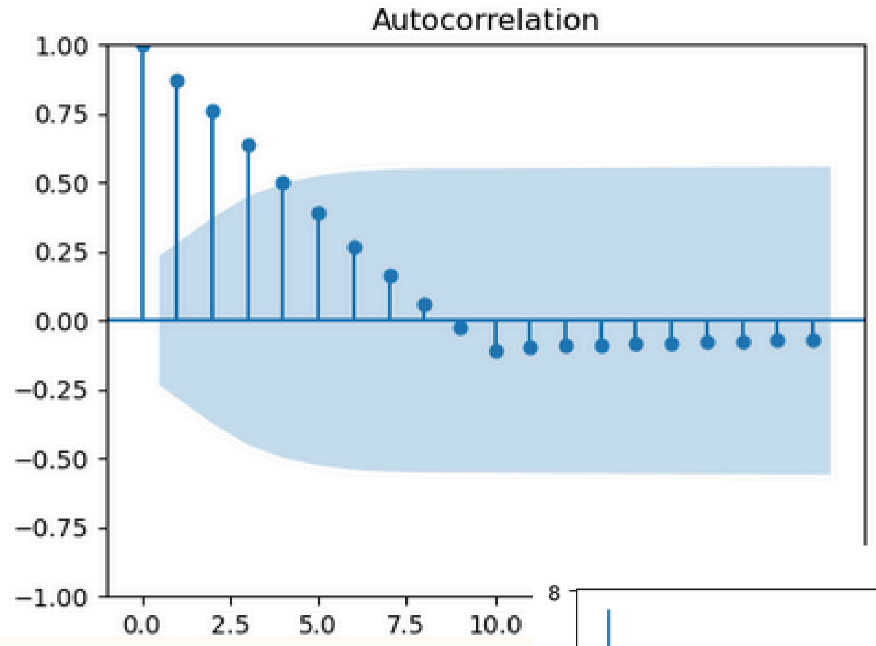
ARIMA Model to forecast Very High Risk Countries' Lack of Coping Capacity Risk indices:

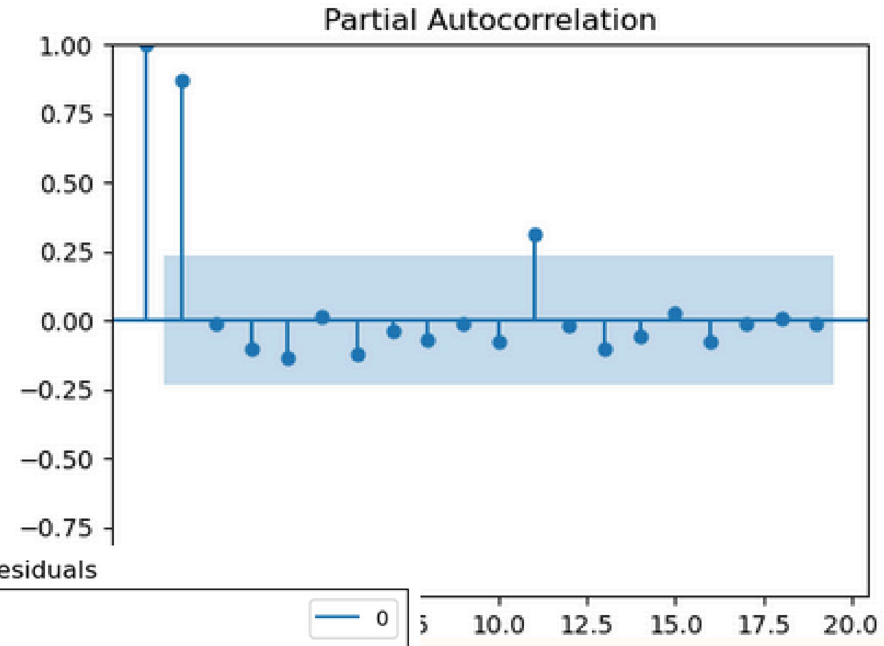
Back to Agenda

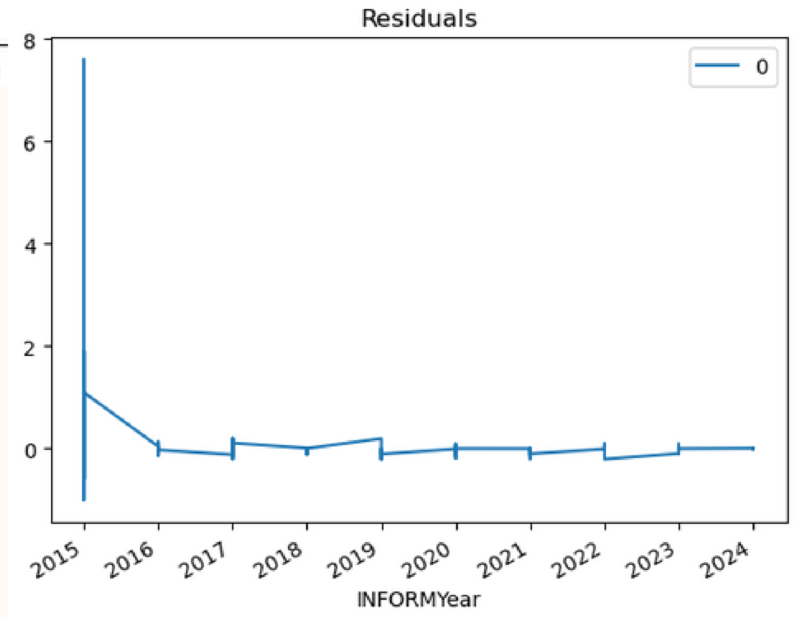










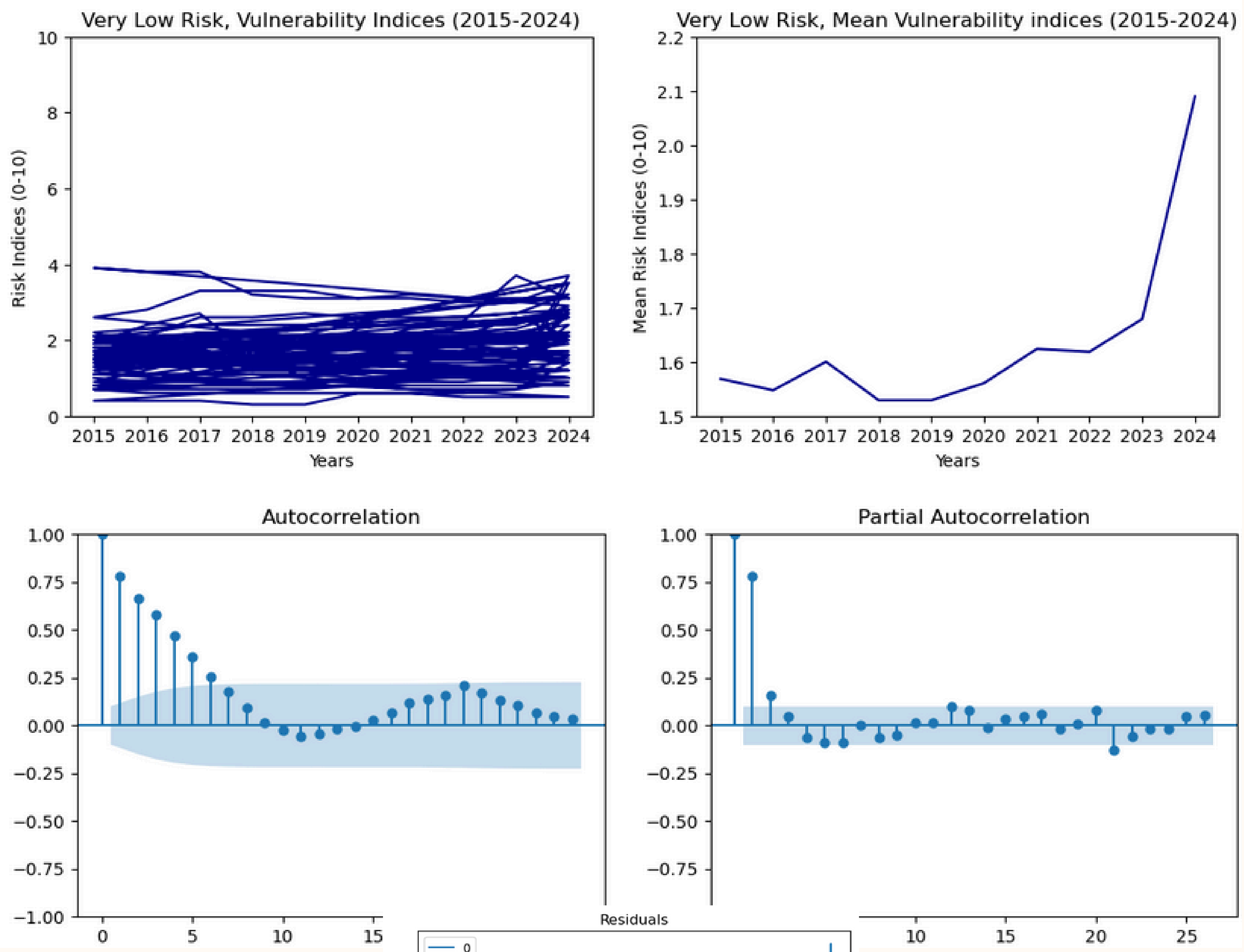


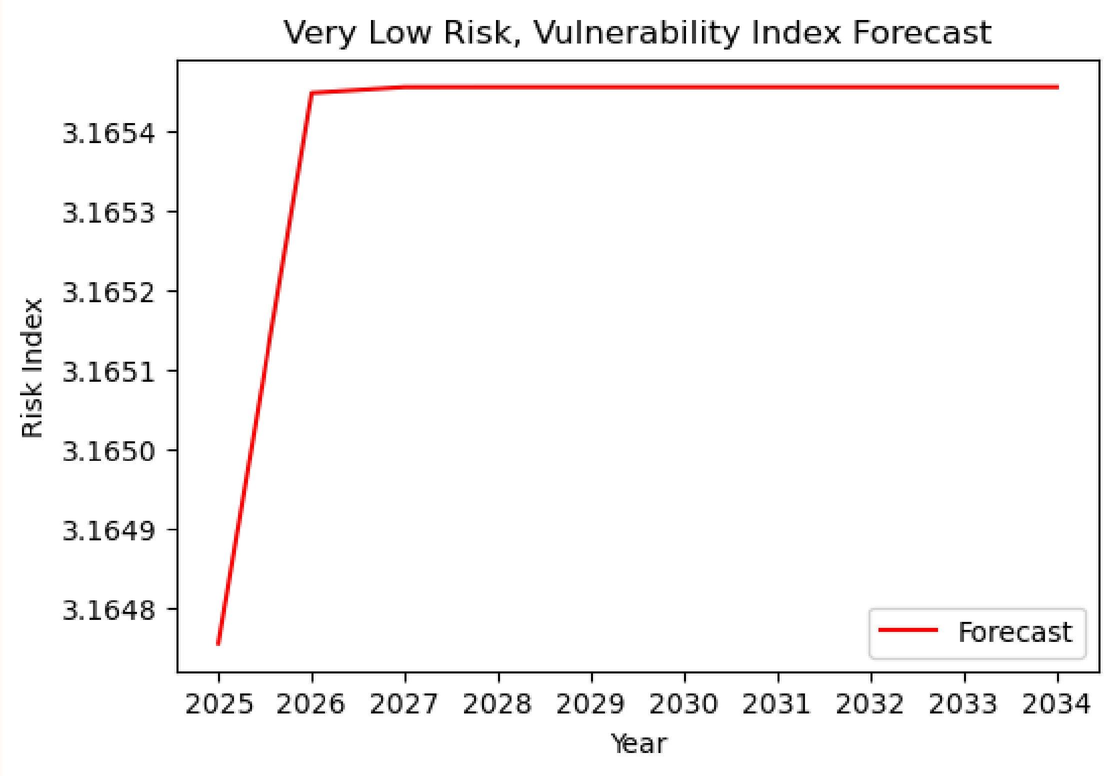

```
Mean Squared Error (MSE): 0.009000015210840626
Mean Absolute Error (MAE): 0.05000048846028733
Root Mean Squared Error (RMSE): 0.09486840997318668
```

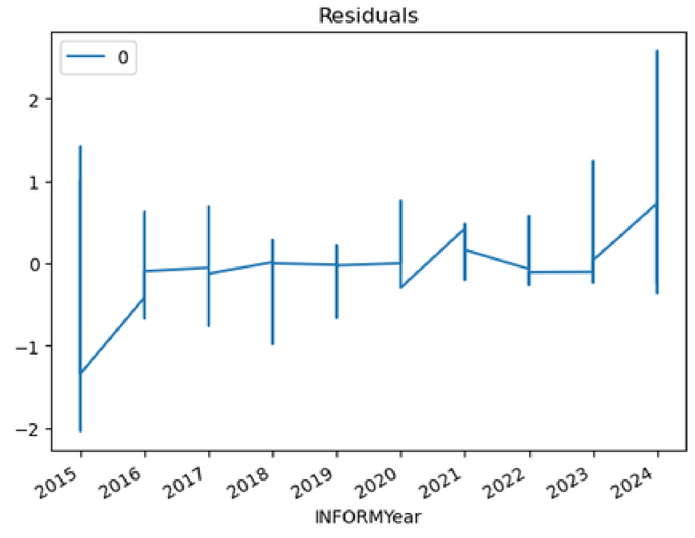
```
                                   SARIMAX Results
==========================================================================================
Dep. Variable:     Lack of Coping Capacity Index   No. Observations:                  70
Model:                            ARIMA(1, 1, 1)   Log Likelihood                -20.815
Date:                           Mon, 08 Apr 2024   AIC                            47.630
Time:                                   04:31:05   BIC                            54.333
Sample:                                        0   HQIC                           50.290
                                            - 70
Covariance Type:                             opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.1428      4.400     -0.032      0.974      -8.767       8.482
ma.L1          0.0866      4.503      0.019      0.985      -8.739       8.913
sigma2         0.1070      0.007     15.921      0.000       0.094       0.120
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):               929.51
Prob(Q):                              1.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.32   Skew:                             2.95
Prob(H) (two-sided):                  0.01   Kurtosis:                        19.98
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

ARIMA Model to forecast Very Low Risk Countries' Vulnerability Risk indices:

ck to Agenda



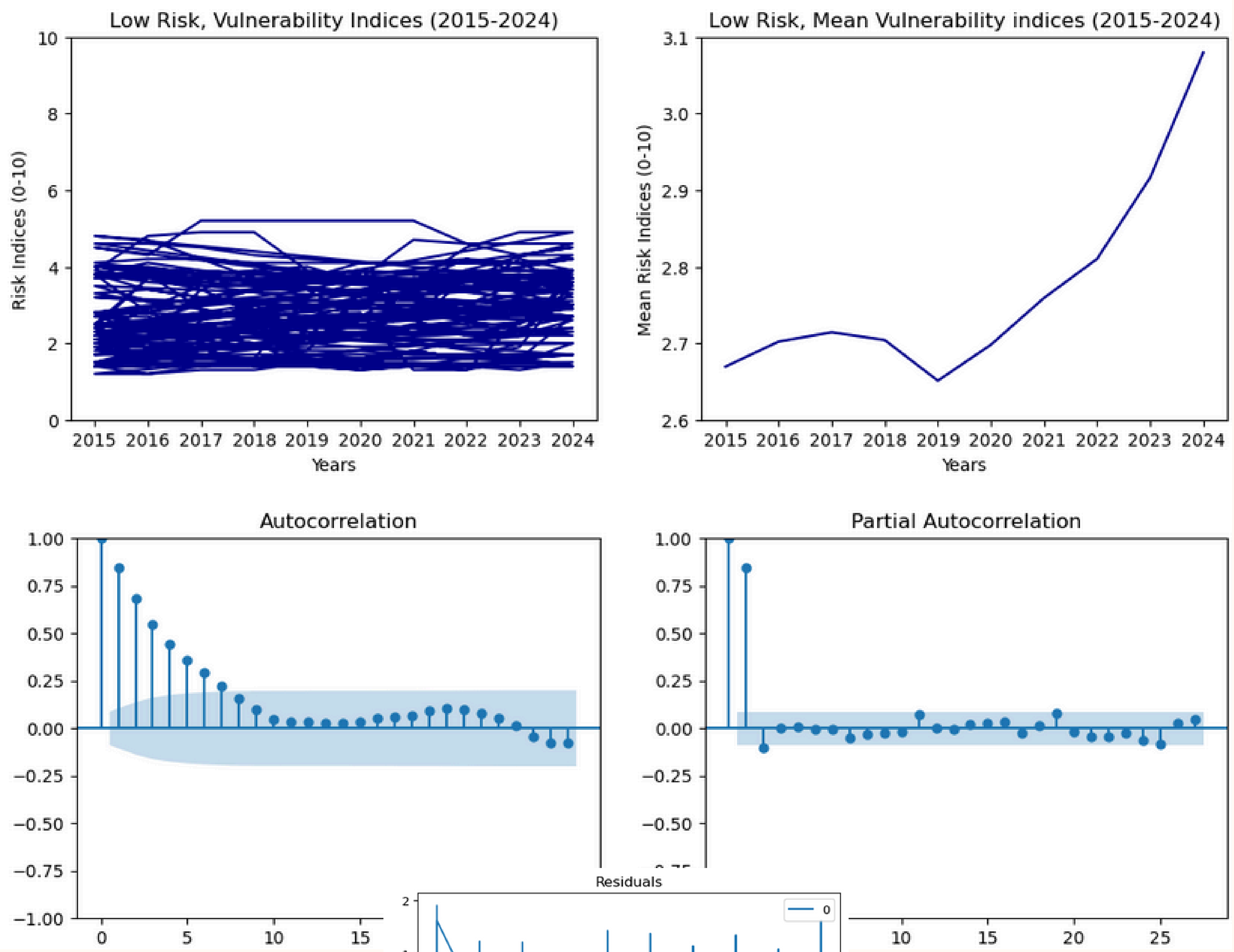


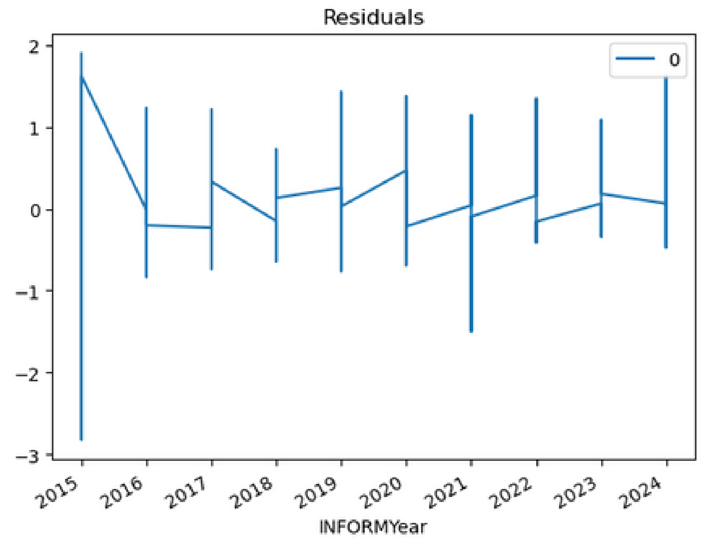


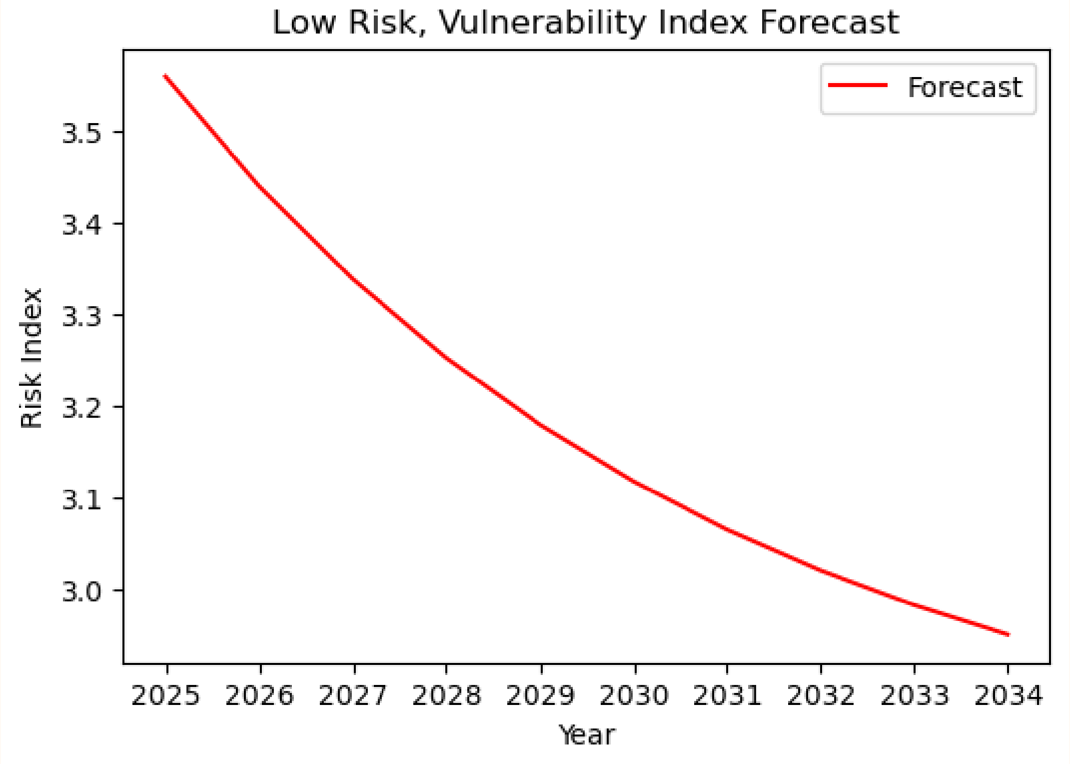

```
Mean Squared Error (MSE): 1.583103628619634
Mean Absolute Error (MAE): 1.1122929107320618
Root Mean Squared Error (RMSE): 1.2582144605033094
                               SARIMAX Results
==============================================================================
Dep. Variable:     Vulnerability Index   No. Observations:                  380
Model:                  ARIMA(1, 1, 1)   Log Likelihood                -223.279
Date:                 Mon, 08 Apr 2024   AIC                            452.558
Time:                         04:36:06   BIC                            464.371
Sample:                              0   HQIC                           457.246
                                 - 380
Covariance Type:                   opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0107      0.202      0.053      0.958      -0.385       0.406
ma.L1         -0.3079      0.205     -1.505      0.132      -0.709       0.093
sigma2         0.1902      0.006     29.751      0.000       0.178       0.203
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              1404.91
Prob(Q):                              0.98   Prob(JB):                         0.00
Heteroskedasticity (H):               1.22   Skew:                             0.40
Prob(H) (two-sided):                  0.27   Kurtosis:                        12.40
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

ARIMA Model to forecast Low Risk Countries' Vulnerability Risk indices:

ack to Agenda





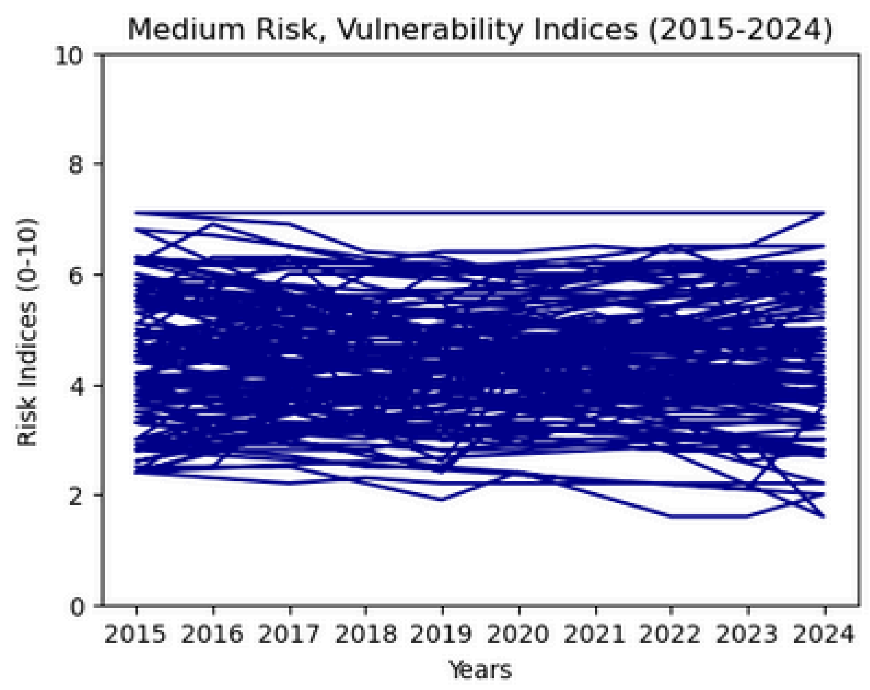


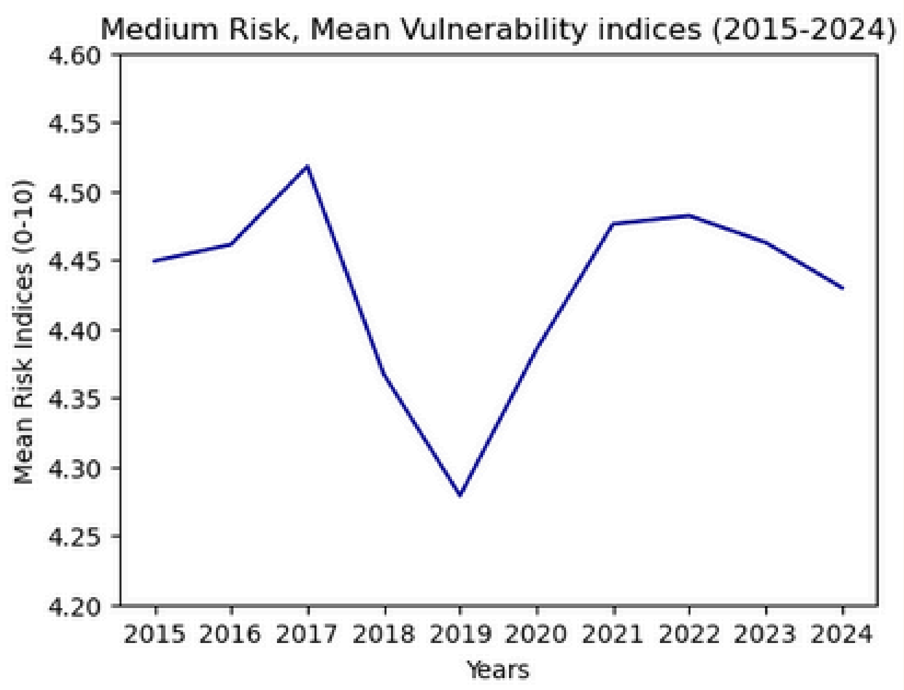

```
Mean Squared Error (MSE): 0.2795727269637688
Mean Absolute Error (MAE): 0.4472900082770751
Root Mean Squared Error (RMSE): 0.5287463730029444
```

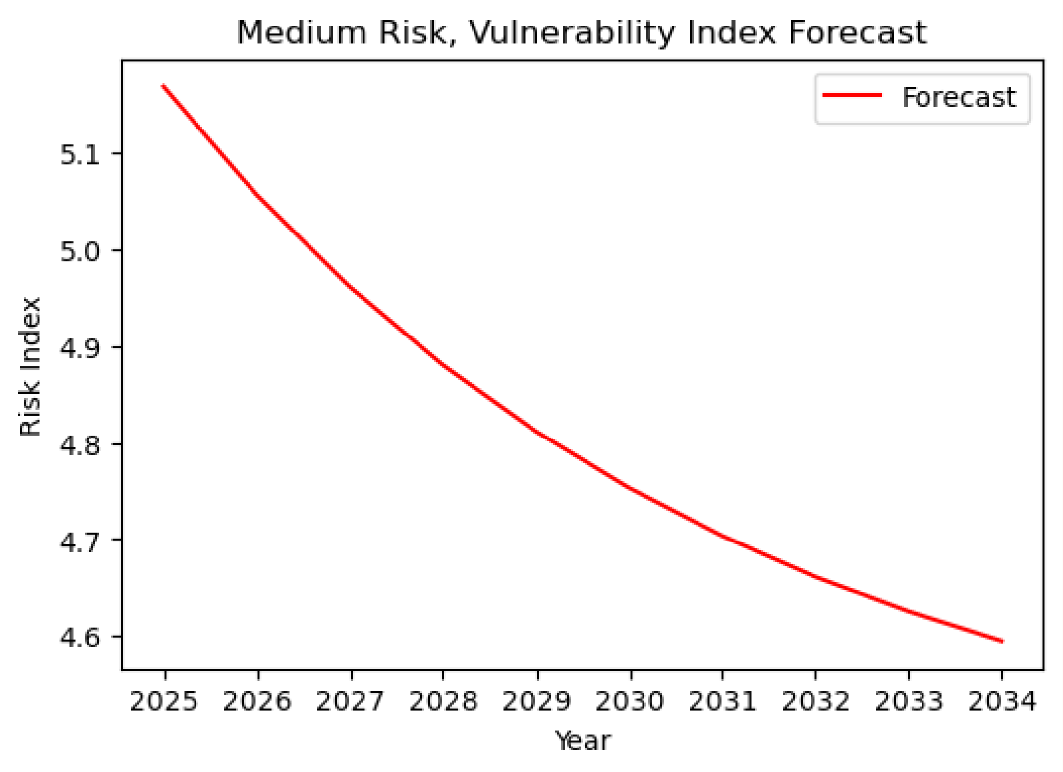
```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Vulnerability Index   No. Observations:                  490
Model:                  ARIMA(1, 1, 1)   Log Likelihood                -366.625
Date:                 Mon, 08 Apr 2024   AIC                            739.249
Time:                         04:39:28   BIC                            751.826
Sample:                              0   HQIC                           744.189
                                 - 490
Covariance Type:                   opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8486      0.032     26.797      0.000       0.787       0.911
ma.L1         -1.0000      3.983     -0.251      0.802      -8.806       6.806
sigma2         0.2603      1.037      0.251      0.802      -1.773       2.294
===================================================================================
Ljung-Box (L1) (Q):                   3.27   Jarque-Bera (JB):               725.70
Prob(Q):                              0.07   Prob(JB):                         0.00
Heteroskedasticity (H):               1.39   Skew:                            -0.58
Prob(H) (two-sided):                  0.03   Kurtosis:                         8.85
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

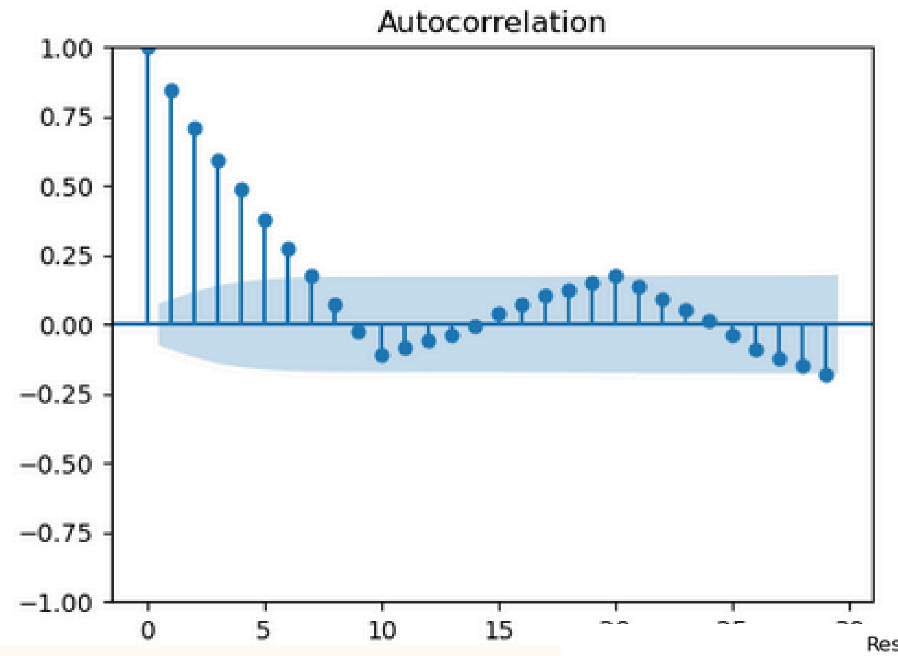
ARIMA Model to forecast Medium Risk Countries’ Vulnerability Risk indices:

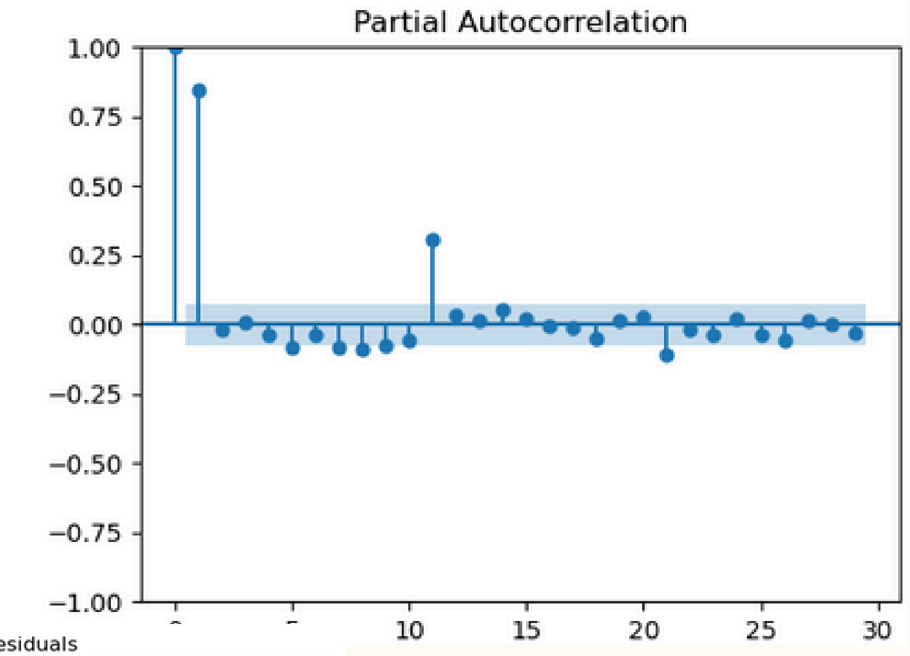
Back to Agenda

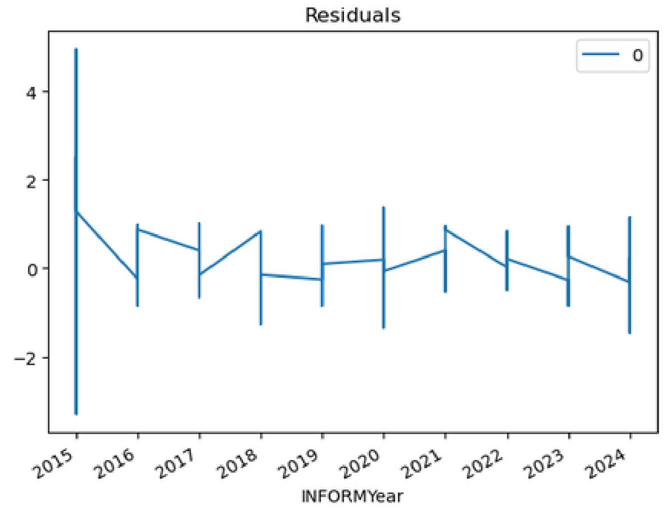








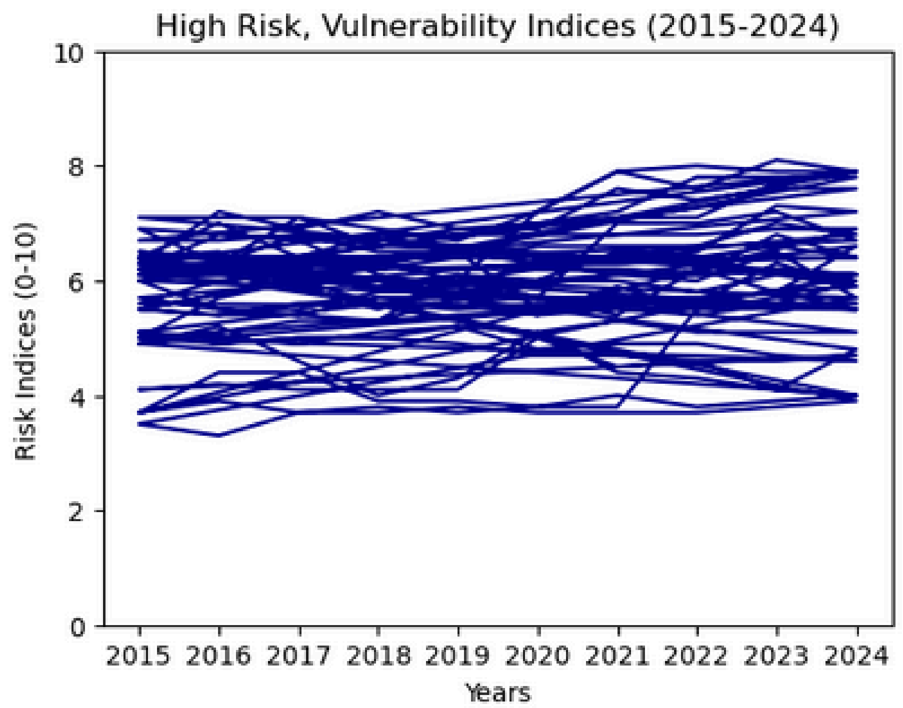


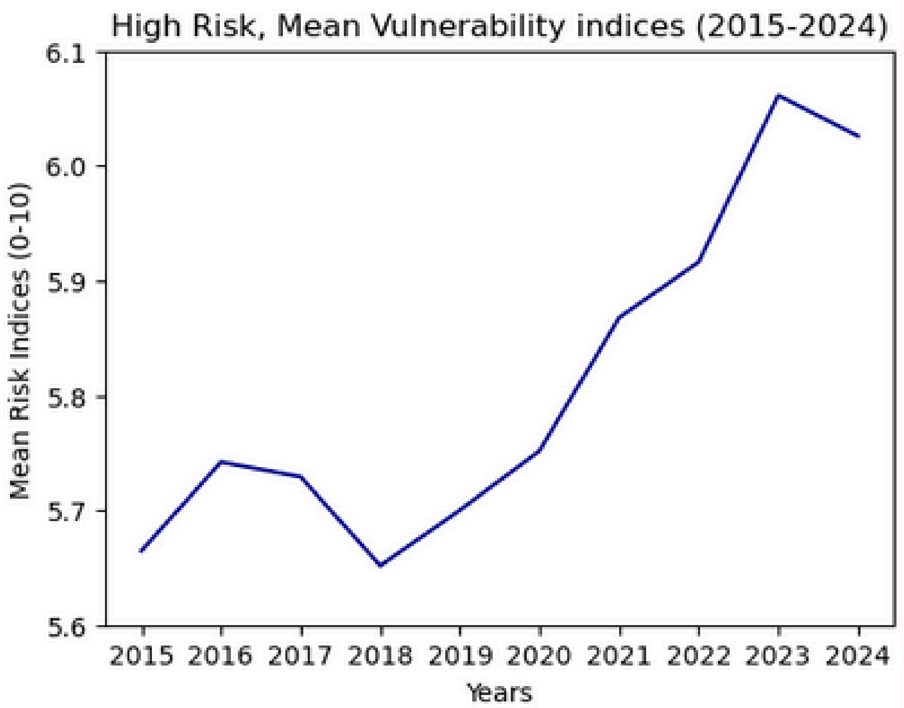


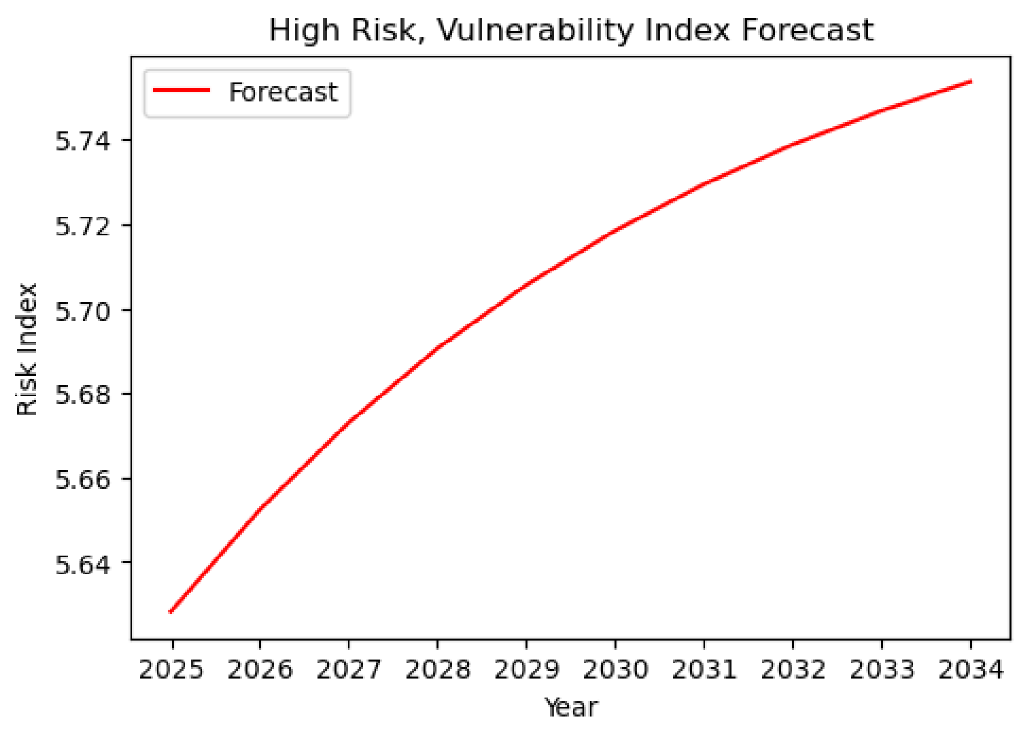

```
Mean Squared Error (MSE): 0.8217599681280845
Mean Absolute Error (MAE): 0.7820684905056288
Root Mean Squared Error (RMSE): 0.906509772770313
                               SARIMAX Results
==============================================================================
Dep. Variable:     Vulnerability Index   No. Observations:                  670
Model:                  ARIMA(1, 1, 1)   Log Likelihood                -633.906
Date:                 Mon, 08 Apr 2024   AIC                           1273.813
Time:                         04:43:45   BIC                           1287.330
Sample:                              0   HQIC                          1279.049
                                 - 670
Covariance Type:                   opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8499      0.030     28.063      0.000       0.791       0.909
ma.L1         -0.9998      0.073    -13.789      0.000      -1.142      -0.858
sigma2         0.3873      0.030     12.873      0.000       0.328       0.446
===================================================================================
Ljung-Box (L1) (Q):                   0.05   Jarque-Bera (JB):              4187.41
Prob(Q):                              0.82   Prob(JB):                         0.00
Heteroskedasticity (H):               1.31   Skew:                             0.56
Prob(H) (two-sided):                  0.05   Kurtosis:                        15.21
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step)
```

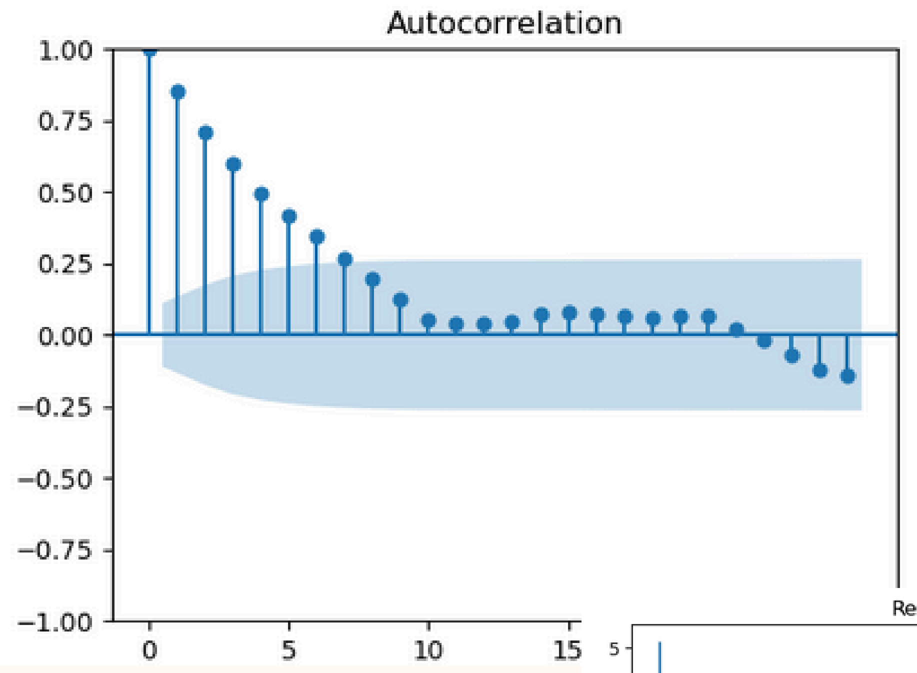
ARIMA Model to forecast High Risk Countries' Vulnerability Risk indices:

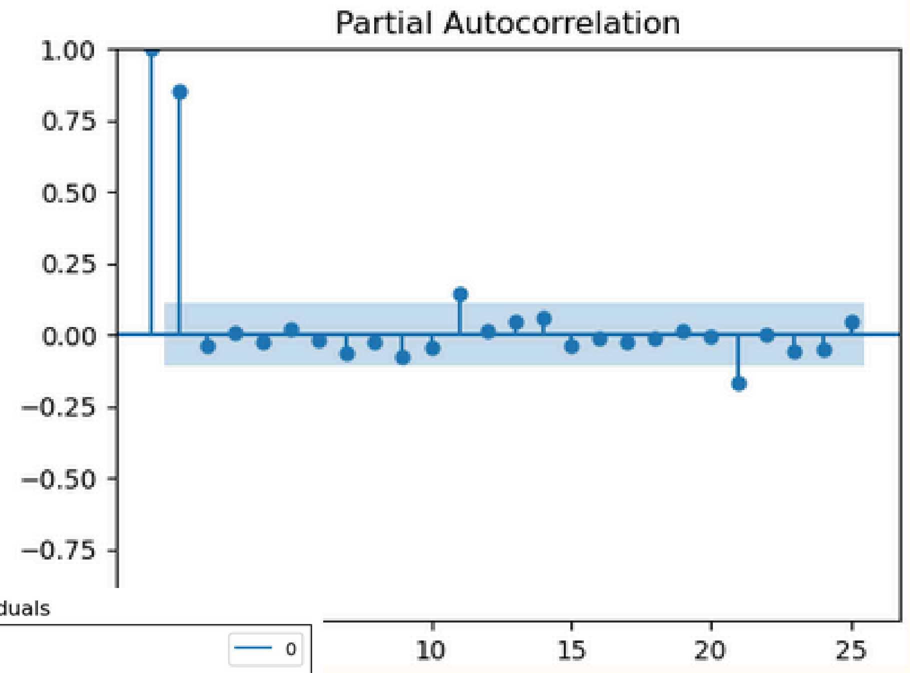
ack to Agenda

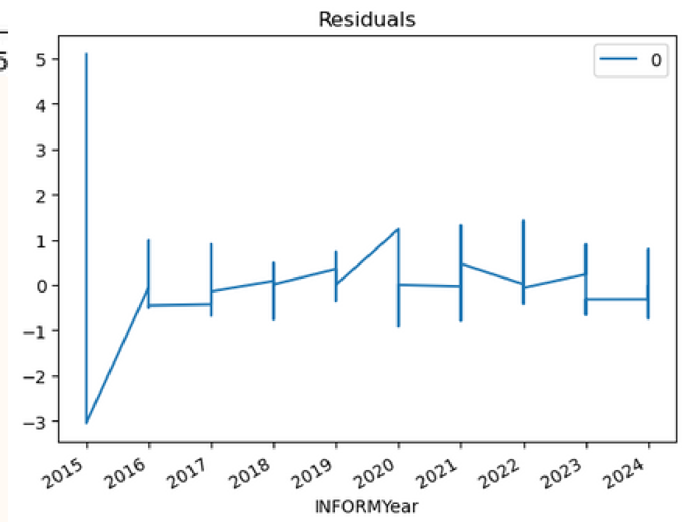










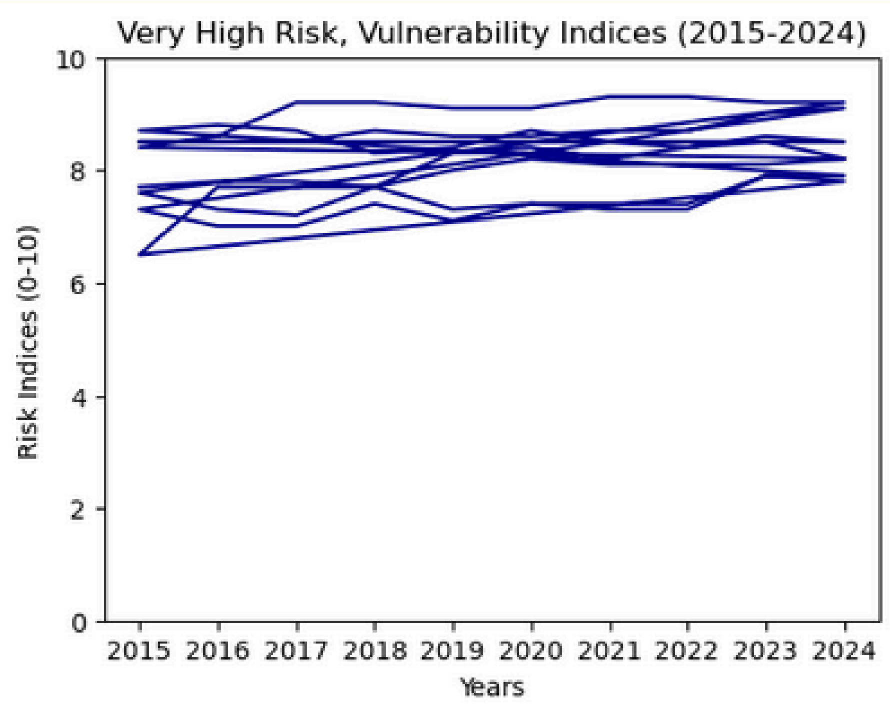


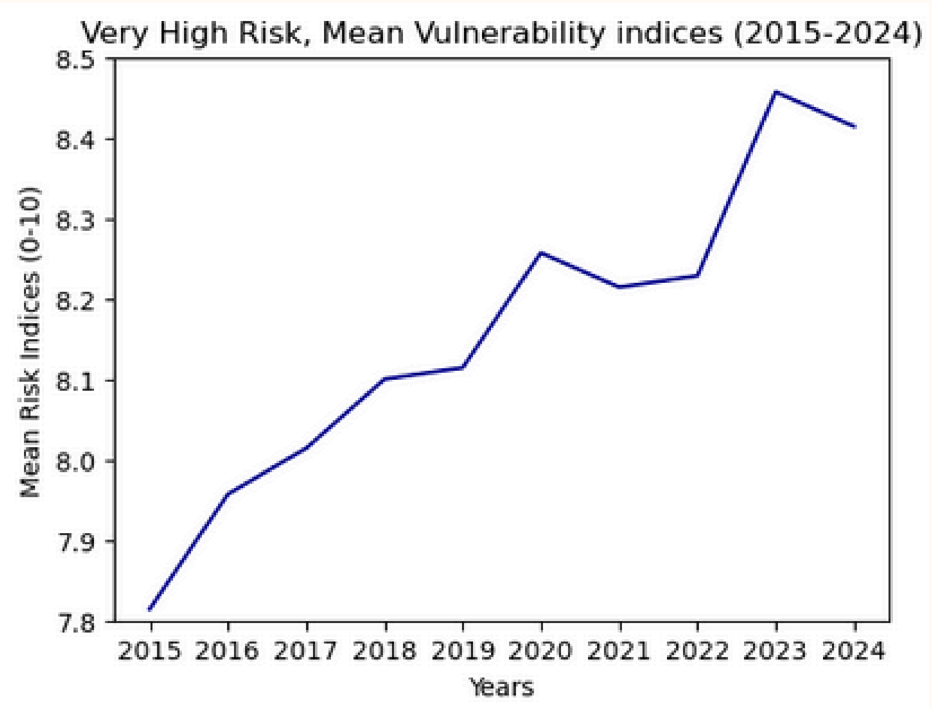

```
Mean Squared Error (MSE): 2.087772772142515
Mean Absolute Error (MAE): 1.243671031554546
Root Mean Squared Error (RMSE): 1.444912721288907
```

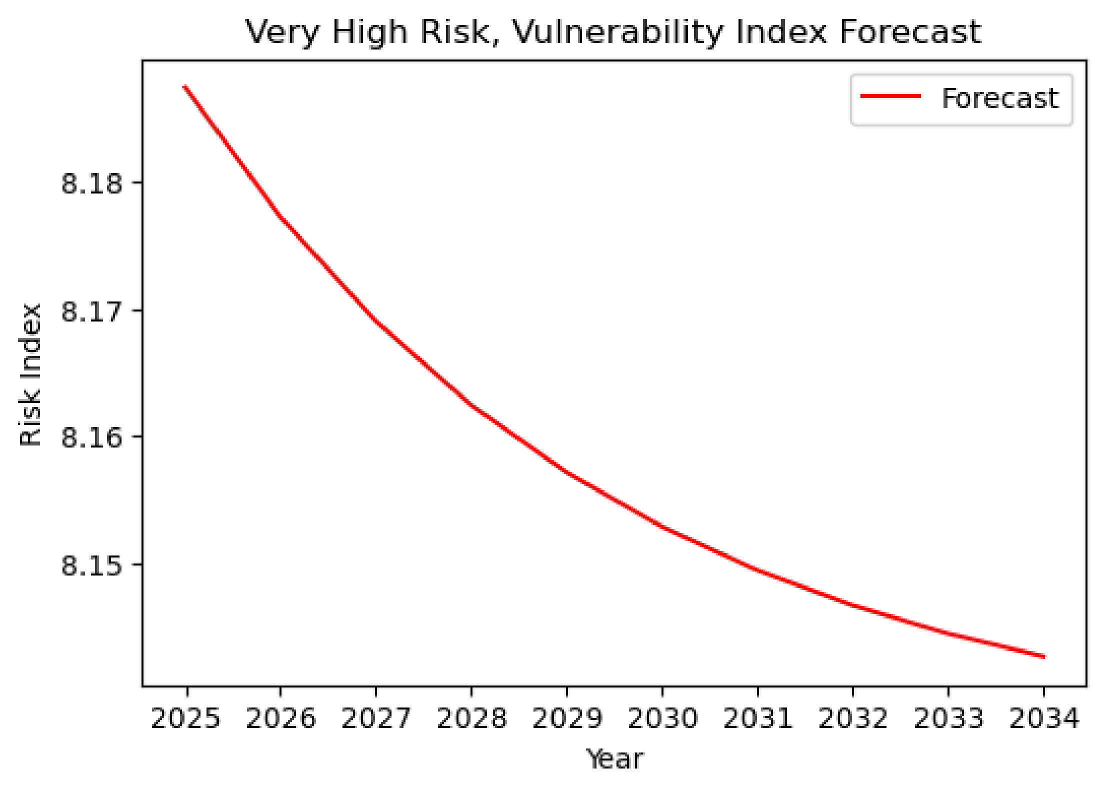
```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Vulnerability Index   No. Observations:                  310
Model:                  ARIMA(1, 1, 1)   Log Likelihood                -250.370
Date:                 Mon, 08 Apr 2024   AIC                            506.741
Time:                         04:46:43   BIC                            517.941
Sample:                              0   HQIC                           511.218
                                 - 310
Covariance Type:                   opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8547      0.036     24.018      0.000       0.785       0.924
ma.L1         -1.0000      2.451     -0.408      0.683      -5.804       3.804
sigma2         0.2929      0.722      0.406      0.685      -1.121       1.707
===================================================================================
Ljung-Box (L1) (Q):                   0.27   Jarque-Bera (JB):               718.05
Prob(Q):                              0.60   Prob(JB):                         0.00
Heteroskedasticity (H):               1.06   Skew:                            -0.83
Prob(H) (two-sided):                  0.77   Kurtosis:                        10.28
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

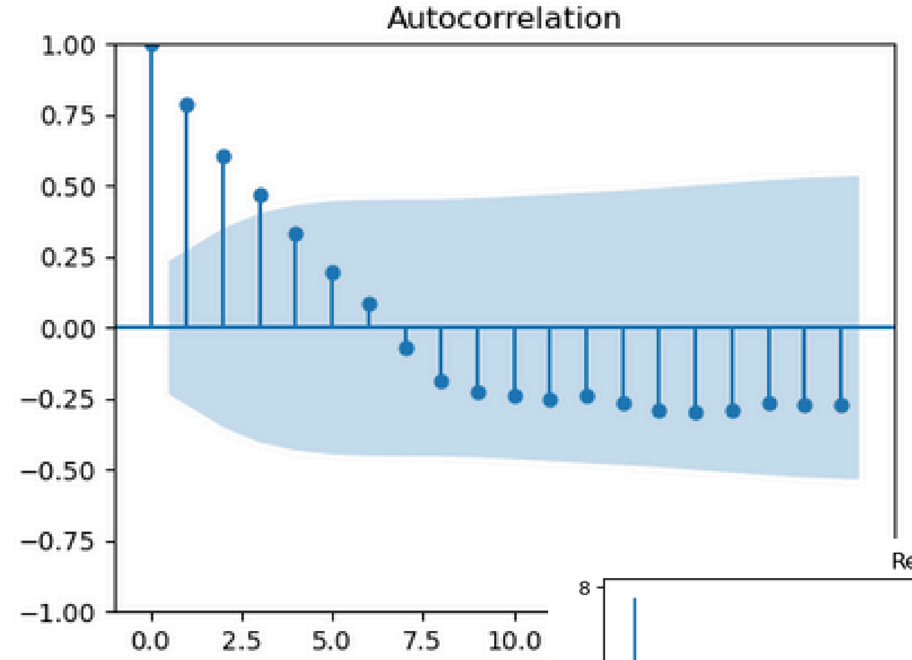
ARIMA Model to forecast Very High Risk Countries' Vulnerability Risk indices:

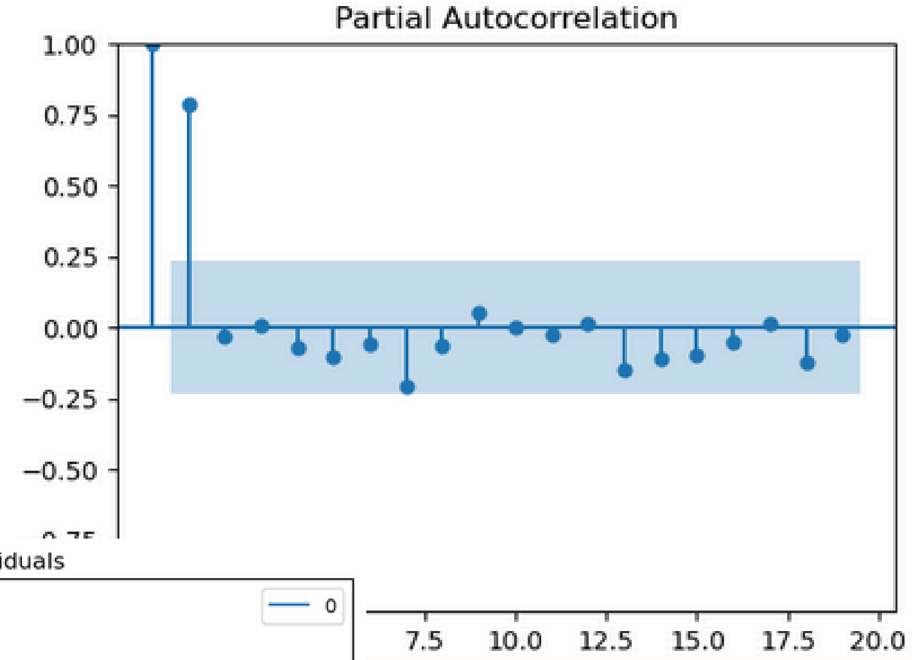
...ck to Agenda

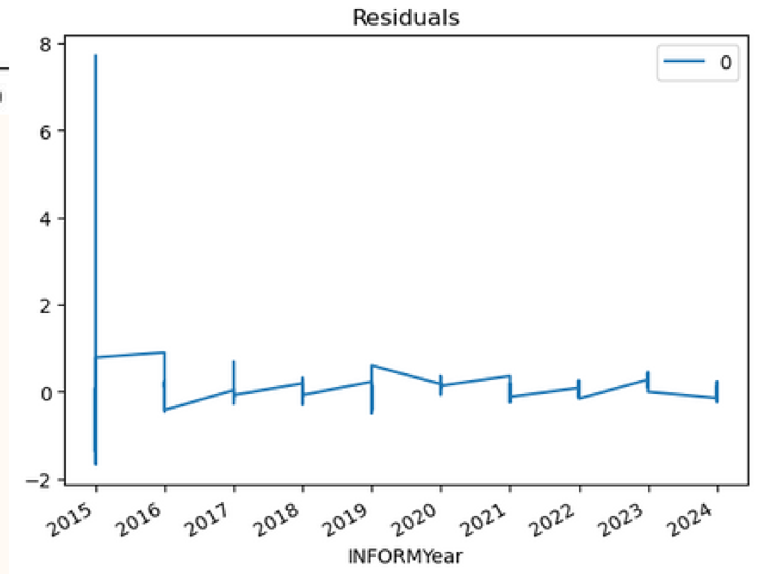












Mean Squared Error (MSE): 0.3545506871275879
Mean Absolute Error (MAE): 0.34981788796145186
Root Mean Squared Error (RMSE): 0.5954415900217148

```
                               SARIMAX Results
==============================================================================
Dep. Variable:     Vulnerability Index   No. Observations:                   70
Model:                  ARIMA(1, 1, 1)   Log Likelihood                 -37.413
Date:                 Mon, 08 Apr 2024   AIC                             80.826
Time:                         04:49:50   BIC                             87.528
Sample:                              0   HQIC                            83.485
                                  - 70
Covariance Type:                   opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8045      0.133      6.039      0.000       0.543       1.066
ma.L1         -1.0000    206.818     -0.005      0.996    -406.356     404.356
sigma2         0.1679     34.716      0.005      0.996     -67.873      68.209
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):               128.25
Prob(Q):                              0.92   Prob(JB):                         0.00
Heteroskedasticity (H):               1.52   Skew:                            -1.80
Prob(H) (two-sided):                  0.32   Kurtosis:                         8.62
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step)
```