In [19]:
```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import numpy as np
```

In [2]:
```
Datasetraw = pd.read_csv("DATA SET.csv")
Datasetraw.head()
```

Out[2]:

| | CLIMATE CHANGE DATA | Unnamed: 1 |
|---|---|---|
| **0** | NaN | Climate |
| **1** | NaN | Change( |
| **2** | NaN | degree |
| **3** | Year | celcius) |
| **4** | 1992 | 0.22 |

In [3]:
```
Datasetraw.sample(10)
```

Out[3]:

| | CLIMATE CHANGE DATA | Unnamed: 1 |
|---|---|---|
| **39** | 1996 | 0.491 |
| **91** | 2013 | 1.454 |
| **29** | 2017 | 0.93 |
| **140** | 1997 | 38.64 |
| **136** | 1993 | 32.95 |
| **66** | NaN | (cubic |
| **138** | 1995 | 39.54 |
| **11** | 1999 | 0.39 |
| **153** | 2010 | 31.06 |
| **3** | Year | celcius) |

Csv file contains data for different Time Series and each would need to be extracted and worked on individually

# CLIMATE CHANGE DATA AND FORECAST

## Etracting The Climate change Data

In [4]:
```
Data_Climate_Change = Datasetraw.iloc[4 : 32]
Data_Climate_Change.columns =["Year", "Climate Change( degree celcius)"]
Data_Climate_Change["Year"]= pd.to_datetime(Data_Climate_Change["Year"])
Data_Climate_Change.set_index("Year", inplace=True)
Data_Climate_Change.head()
```

```
C:\Users\Adamu Yusuf\Anaconda3\lib\site-packages\ipykernel_launcher.py:3: Setti
ngWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stab
le/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas-doc
s/stable/indexing.html#indexing-view-versus-copy)
  This is separate from the ipykernel package so we can avoid doing imports unt
il
```

Out[4]:

| Year | Climate Change( degree celcius) |
|---|---|
| **1992-01-01** | 0.22 |
| **1993-01-01** | 0.23 |
| **1994-01-01** | 0.32 |
| **1995-01-01** | 0.45 |
| **1996-01-01** | 0.33 |

In [ ]:
```

```

In [5]:
```
Data_Climate_Change['Climate Change( degree celcius)'] = pd.to_numeric(Data_
Data_Climate_Change.shape

```

```
C:\Users\Adamu Yusuf\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: Setti
ngWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stab
le/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas-doc
s/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.
```

Out[5]: (28, 1)

In [6]:
```
1 Data_Climate_Change.describe()
```

Out[6]:

| | Climate Change( degree celcius) |
|---|---|
| **count** | 28.000000 |
| **mean** | 0.609286 |
| **std** | 0.212183 |
| **min** | 0.220000 |
| **25%** | 0.465000 |
| **50%** | 0.625000 |
| **75%** | 0.700000 |
| **max** | 1.020000 |

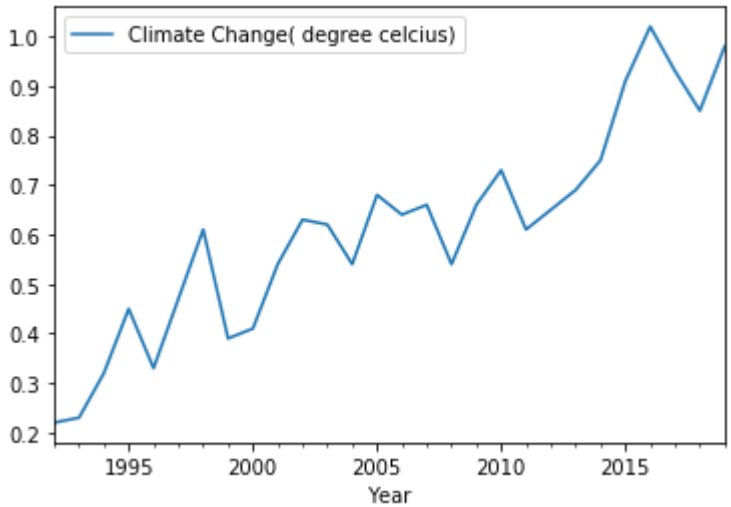
In [7]:
```
1 Data_Climate_Change.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 28 entries, 1992-01-01 to 2019-01-01
Data columns (total 1 columns):
Climate Change( degree celcius)    28 non-null float64
dtypes: float64(1)
memory usage: 448.0 bytes
```

## Graph of the Data

In [8]:
```
1 Data_Climate_Change.plot()
```

Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x28dd1f73438>




In [9]:
```
1 #Checking for Stationarilty
2 from statsmodels.tsa.stattools import adfuller
```

In [10]:
```
1 test_results = adfuller(Data_Climate_Change["Climate Change( degree celcius)
```

In [11]:
```
1 test_results
```

Out[11]:
```
(-1.402874006275917,
 0.5808975108286634,
 0,
 27,
 {'1%': -3.6996079738860943,
  '5%': -2.9764303469999494,
  '10%': -2.627601001371742},
 -32.29182166798443)
```

Data is not Stationary and needs to undergo transformation

In [12]:
```
1 #Transforming the data using 1at dfference
2 Data_Climate_Change["Change difference"] = Data_Climate_Change["Climate Chan
```

```
C:\Users\Adamu Yusuf\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: Setti
ngWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stab
le/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas-doc
s/stable/indexing.html#indexing-view-versus-copy)
```

In [13]:
```
1 Data_Climate_Change.head()
```

Out[13]:

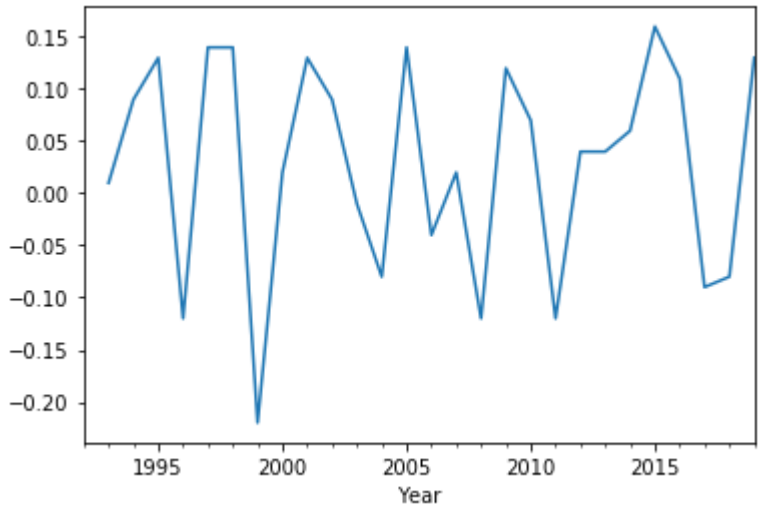
| Year | Climate Change( degree celcius) | Change difference |
|---|---|---|
| **1992-01-01** | 0.22 | NaN |
| **1993-01-01** | 0.23 | 0.01 |
| **1994-01-01** | 0.32 | 0.09 |
| **1995-01-01** | 0.45 | 0.13 |
| **1996-01-01** | 0.33 | -0.12 |

In [14]:
```
1 test_results = adfuller(Data_Climate_Change["Change difference"].dropna())
2 test_results
```

Out[14]:
```
(-6.4498533557201885,
 1.5330920059031378e-08,
 1,
 25,
 {'1%': -3.7238633119999998, '5%': -2.98648896, '10%': -2.6328004},
 -30.166533709716333)
```

In [15]:
```
1 Data_Climate_Change["Change difference"].plot()
```

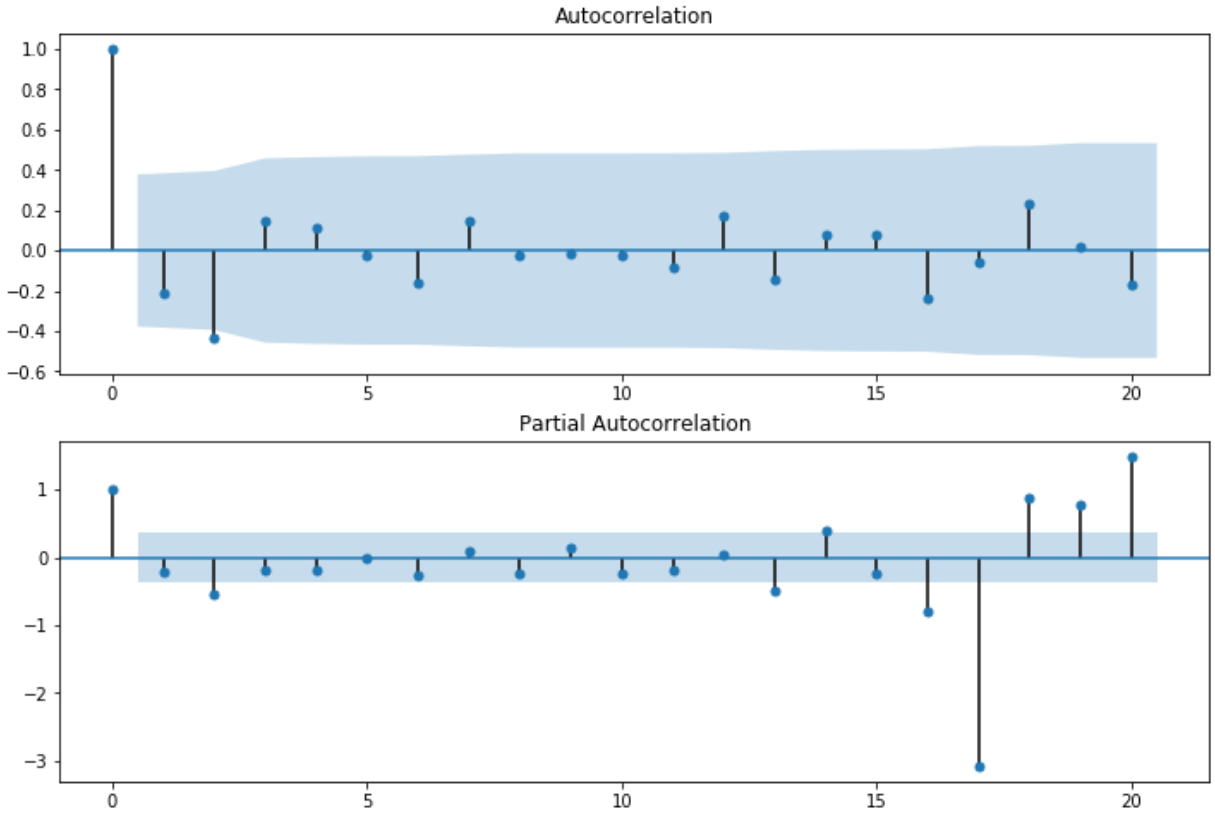
Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x28dd40dd6a0>




As can be seen from the above graph and adfuller test result Data has been succesfuly transformed to a stationary data

In [16]:
```
1 from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
```

In [24]:
```
fig = plt.figure(figsize=(12,8))
ax1 = fig.add_subplot(211)
fig = plot_acf(Data_Climate_Change["Change difference"].dropna(), lags = 20,
ax2 = fig.add_subplot(212)
fig = plot_pacf(Data_Climate_Change["Change difference"].dropna(), lags = 20
```




In [ ]:
```
#from pacf and acf graph respectively (p = 5, q= 5)
#d =1 (first difference)

```

## MODEL-1 ARIMA

In [82]:
```
from statsmodels.tsa.arima_model import ARIMA
model = ARIMA(Data_Climate_Change["Climate Change( degree celcius)"], order=
model_fit = model.fit()
```

```
C:\Users\Adamu Yusuf\Anaconda3\lib\site-packages\statsmodels\base\model.py:492:
HessianInversionWarning: Inverting hessian failed, no bse or cov_params availab
le
  'available', HessianInversionWarning)
```

In [83]:
```
1 model_fit.summary()
```

```
C:\Users\Adamu Yusuf\Anaconda3\lib\site-packages\statsmodels\tsa\arima_model.p
y:1441: RuntimeWarning: invalid value encountered in sqrt
  return np.sqrt(np.diag(-inv(hess)))
C:\Users\Adamu Yusuf\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastruct
ure.py:877: RuntimeWarning: invalid value encountered in greater
  return (self.a < x) & (x < self.b)
C:\Users\Adamu Yusuf\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastruct
ure.py:877: RuntimeWarning: invalid value encountered in less
  return (self.a < x) & (x < self.b)
C:\Users\Adamu Yusuf\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastruct
ure.py:1831: RuntimeWarning: invalid value encountered in less_equal
  cond2 = cond0 & (x <= self.a)
```

Out[83]:

ARIMA Model Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | D.Climate Change( degree celcius) | **No. Observations:** | 27 |
| **Model:** | ARIMA(5, 1, 3) | **Log Likelihood** | 30.903 |
| **Method:** | css-mle | **S.D. of innovations** | 0.069 |
| **Date:** | Tue, 10 Nov 2020 | **AIC** | -41.805 |
| **Time:** | 21:48:12 | **BIC** | -28.847 |
| **Sample:** | 01-01-1993 | **HQIC** | -37.952 |
| | - 01-01-2019 | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 0.0241 | 0.003 | 8.898 | 0.000 | 0.019 | 0.029 |
| **ar.L1.D.Climate Change( degree celcius)** | -1.1528 | nan | nan | nan | nan | nan |
| **ar.L2.D.Climate Change( degree celcius)** | -0.4510 | nan | nan | nan | nan | nan |
| **ar.L3.D.Climate Change( degree celcius)** | 0.1296 | nan | nan | nan | nan | nan |
| **ar.L4.D.Climate Change( degree celcius)** | 0.1785 | nan | nan | nan | nan | nan |
| **ar.L5.D.Climate Change( degree celcius)** | 0.3786 | nan | nan | nan | nan | nan |
| **ma.L1.D.Climate Change( degree celcius)** | 0.7424 | 0.097 | 7.619 | 0.000 | 0.551 | 0.933 |
| **ma.L2.D.Climate Change( degree celcius)** | -0.7424 | 0.181 | -4.111 | 0.001 | -1.096 | -0.388 |
| **ma.L3.D.Climate Change( degree celcius)** | -1.0000 | 0.096 | -10.369 | 0.000 | -1.189 | -0.811 |

Roots

| | Real | Imaginary | Modulus | Frequency |
|---|---|---|---|---|
| **AR.1** | -0.8719 | -0.4896j | 1.0000 | -0.4186 |
| **AR.2** | -0.8719 | +0.4896j | 1.0000 | 0.4186 |
| **AR.3** | 1.4510 | -0.0000j | 1.4510 | -0.0000 |
| **AR.4** | -0.0893 | -1.3463j | 1.3492 | -0.2605 |
| **AR.5** | -0.0893 | +1.3463j | 1.3492 | 0.2605 |

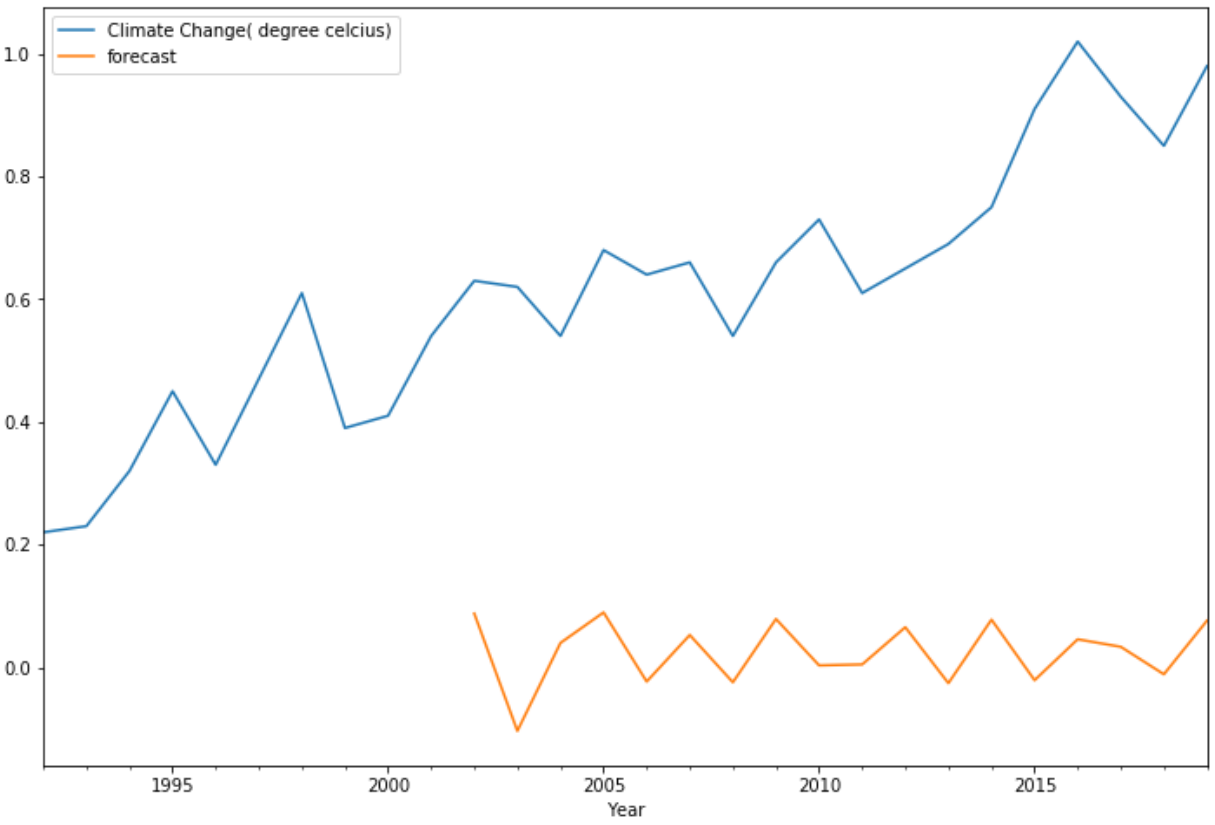
| | | | | |
|---|---|---|---|---|
| **MA.1** | 1.0000 | -0.0000j | 1.0000 | -0.0000 |
| **MA.2** | -0.8712 | -0.4909j | 1.0000 | -0.4183 |
| **MA.3** | -0.8712 | +0.4909j | 1.0000 | 0.4183 |

In [84]:
```
1 Data_Climate_Change["forecast"] = model_fit.predict(start=10,end=27, dynamic
2 Data_Climate_Change[["Climate Change( degree celcius)","forecast"]].plot(fig
```

```
C:\Users\Adamu Yusuf\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: Setti
ngWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stab
le/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas-doc
s/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.
```

Out[84]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1ffefb117b8>
```




**ARIMA performs verypoorly and predictions are far from actual value**

as we can see from the graph, the model performs pooorly on the climate change dataset most likely due to its seasonality.

In [85]:
```
1 Data_Climate_Change
```

Out[85]:

| Year | Climate Change( degree celcius) | Change difference | forecast |
|---|---|---|---|
| **1992-01-01** | 0.22 | NaN | NaN |
| **1993-01-01** | 0.23 | 0.01 | NaN |
| **1994-01-01** | 0.32 | 0.09 | NaN |
| **1995-01-01** | 0.45 | 0.13 | NaN |
| **1996-01-01** | 0.33 | -0.12 | NaN |
| **1997-01-01** | 0.47 | 0.14 | NaN |
| **1998-01-01** | 0.61 | 0.14 | NaN |
| **1999-01-01** | 0.39 | -0.22 | NaN |
| **2000-01-01** | 0.41 | 0.02 | NaN |
| **2001-01-01** | 0.54 | 0.13 | NaN |
| **2002-01-01** | 0.63 | 0.09 | 0.087587 |
| **2003-01-01** | 0.62 | -0.01 | -0.103780 |
| **2004-01-01** | 0.54 | -0.08 | 0.039671 |
| **2005-01-01** | 0.68 | 0.14 | 0.089399 |
| **2006-01-01** | 0.64 | -0.04 | -0.023354 |
| **2007-01-01** | 0.66 | 0.02 | 0.052580 |
| **2008-01-01** | 0.54 | -0.12 | -0.024503 |
| **2009-01-01** | 0.66 | 0.12 | 0.078684 |
| **2010-01-01** | 0.73 | 0.07 | 0.003033 |
| **2011-01-01** | 0.61 | -0.12 | 0.004588 |
| **2012-01-01** | 0.65 | 0.04 | 0.065274 |
| **2013-01-01** | 0.69 | 0.04 | -0.025955 |
| **2014-01-01** | 0.75 | 0.06 | 0.077608 |
| **2015-01-01** | 0.91 | 0.16 | -0.021135 |
| **2016-01-01** | 1.02 | 0.11 | 0.045590 |
| **2017-01-01** | 0.93 | -0.09 | 0.033312 |
| **2018-01-01** | 0.85 | -0.08 | -0.011475 |
| **2019-01-01** | 0.98 | 0.13 | 0.075922 |

In [ ]:
```
1
```

## Model 2 SARIMAX

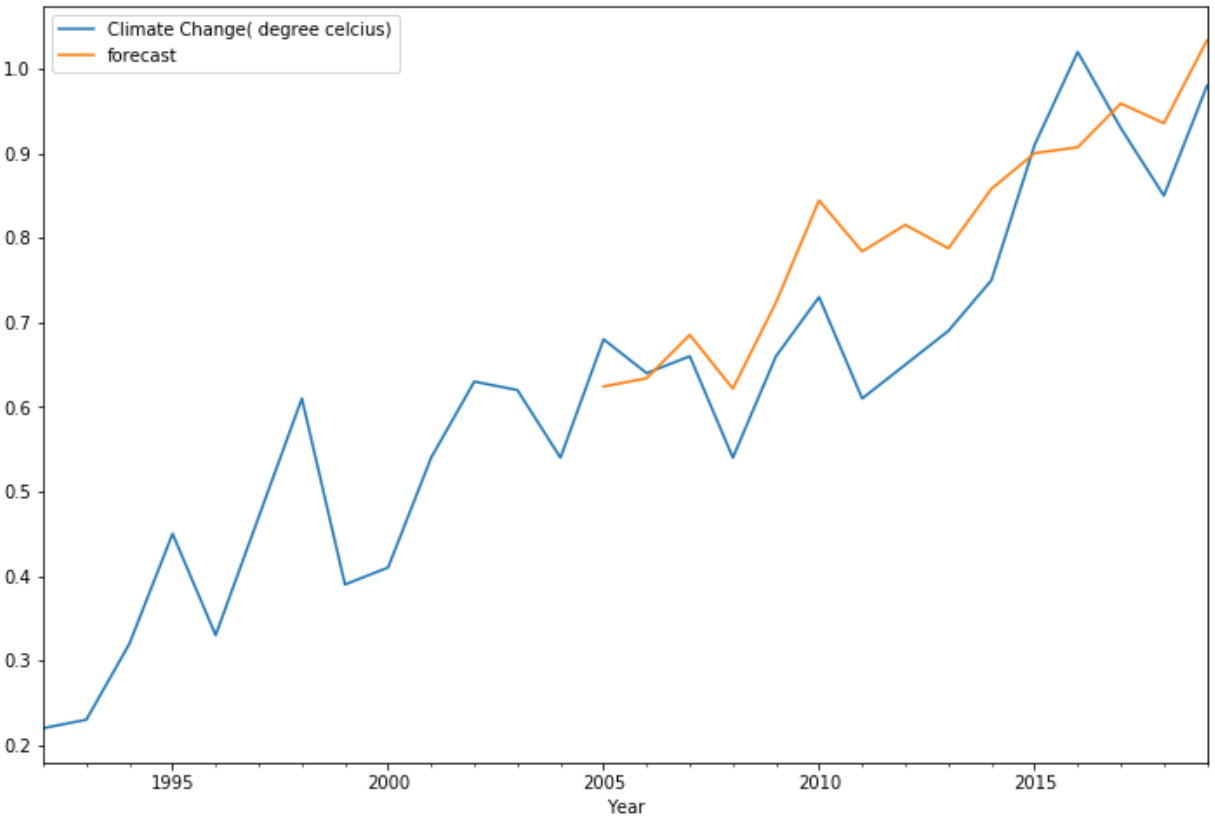
In [110]:
```
import statsmodels.api as sm
model = sm.tsa.statespace.SARIMAX(Data_Climate_Change["Climate Change( degree celci
results = model.fit()
```

In [132]:
```
Data_Climate_Change["forecast"]=results.predict(start=13,end=27,dynamic=True
Data_Climate_Change[["Climate Change( degree celcius)","forecast"]].plot(fig
```

```
C:\Users\Adamu Yusuf\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: Setti
ngWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stab
le/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas-doc
s/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.
```

Out[132]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1fff6ba6a20>
```



In [133]:
```
1 Data_Climate_Change
```

Out[133]:

| Year | Climate Change( degree celcius) | Change difference | forecast |
|---|---|---|---|
| **1992-01-01** | 0.22 | NaN | NaN |
| **1993-01-01** | 0.23 | 0.01 | NaN |
| **1994-01-01** | 0.32 | 0.09 | NaN |
| **1995-01-01** | 0.45 | 0.13 | NaN |
| **1996-01-01** | 0.33 | -0.12 | NaN |
| **1997-01-01** | 0.47 | 0.14 | NaN |
| **1998-01-01** | 0.61 | 0.14 | NaN |
| **1999-01-01** | 0.39 | -0.22 | NaN |
| **2000-01-01** | 0.41 | 0.02 | NaN |
| **2001-01-01** | 0.54 | 0.13 | NaN |
| **2002-01-01** | 0.63 | 0.09 | NaN |
| **2003-01-01** | 0.62 | -0.01 | NaN |
| **2004-01-01** | 0.54 | -0.08 | NaN |
| **2005-01-01** | 0.68 | 0.14 | 0.624154 |
| **2006-01-01** | 0.64 | -0.04 | 0.634013 |
| **2007-01-01** | 0.66 | 0.02 | 0.685279 |
| **2008-01-01** | 0.54 | -0.12 | 0.621625 |
| **2009-01-01** | 0.66 | 0.12 | 0.723604 |
| **2010-01-01** | 0.73 | 0.07 | 0.844374 |
| **2011-01-01** | 0.61 | -0.12 | 0.783980 |
| **2012-01-01** | 0.65 | 0.04 | 0.815496 |
| **2013-01-01** | 0.69 | 0.04 | 0.787651 |
| **2014-01-01** | 0.75 | 0.06 | 0.858143 |
| **2015-01-01** | 0.91 | 0.16 | 0.900194 |
| **2016-01-01** | 1.02 | 0.11 | 0.907228 |
| **2017-01-01** | 0.93 | -0.09 | 0.959062 |
| **2018-01-01** | 0.85 | -0.08 | 0.935740 |
| **2019-01-01** | 0.98 | 0.13 | 1.033646 |

# Predicting for future years

In [174]:
```
from pandas.tseries.offsets import DateOffset
future_dates = [Data_Climate_Change.index[-1] + DateOffset(years=x) for x in
```

In [175]:
```
future_dates_Data_Climate_Change = pd.DataFrame(index=future_dates[1:], colu
```

In [176]:
```
future_dates_Data_Climate_Change.tail()
```

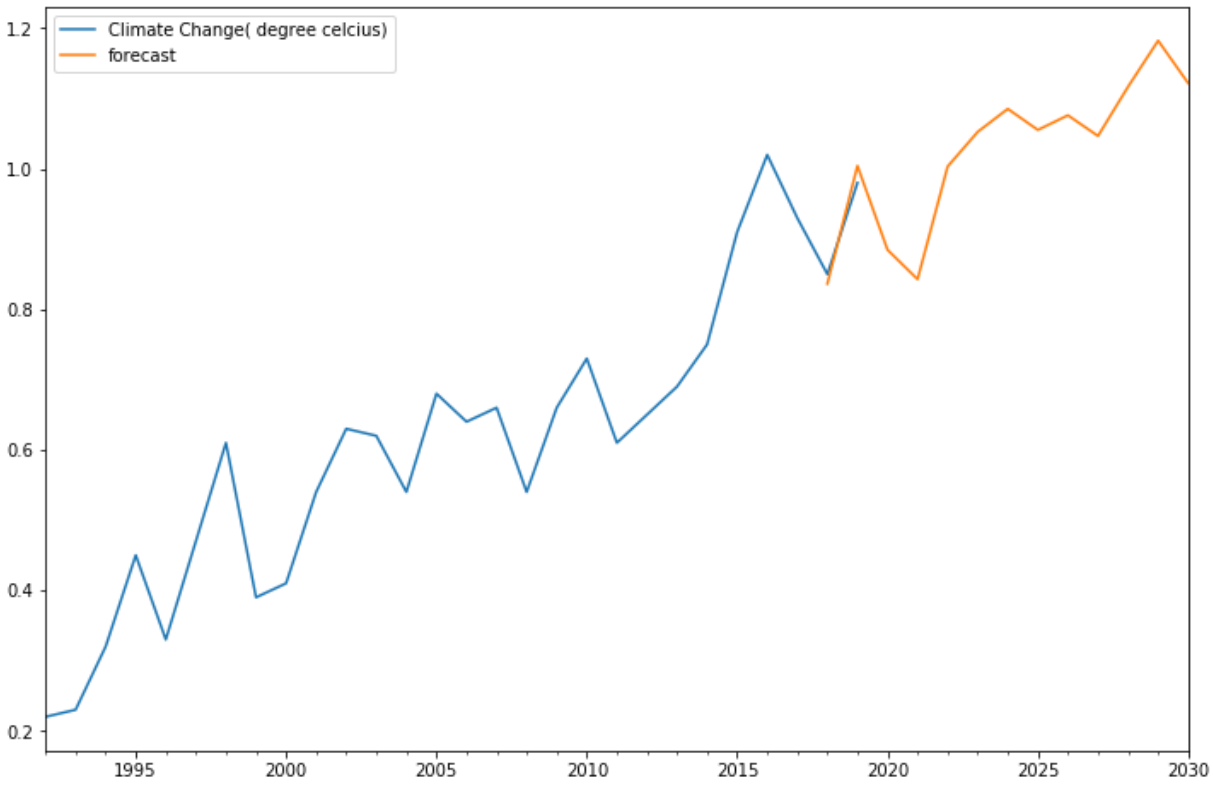
Out[176]:

| | Climate Change( degree celcius) | Change difference | forecast |
|---|---|---|---|
| **2026-01-01** | NaN | NaN | NaN |
| **2027-01-01** | NaN | NaN | NaN |
| **2028-01-01** | NaN | NaN | NaN |
| **2029-01-01** | NaN | NaN | NaN |
| **2030-01-01** | NaN | NaN | NaN |

In [177]:
```
future_dates_Data_Climate_Change = pd.concat([Data_Climate_Change,future_dat
```

In [178]:
```
future_dates_Data_Climate_Change["forecast"] = results.predict(start=26, end
future_dates_Data_Climate_Change[["Climate Change( degree celcius)","forecas
```

Out[178]: <matplotlib.axes._subplots.AxesSubplot at 0x1fff7f834a8>




# ACTUAL DATA(1992-2020) AND FORECAST(2020-2030)

In [179]:
```
future_dates_Data_Climate_Change.drop("Change difference",inplace=True, axis
future_dates_Data_Climate_Change
```

Out[179]:

| | Climate Change( degree celcius) | forecast |
|---|---|---|
| **1992-01-01** | 0.22 | NaN |
| **1993-01-01** | 0.23 | NaN |
| **1994-01-01** | 0.32 | NaN |
| **1995-01-01** | 0.45 | NaN |
| **1996-01-01** | 0.33 | NaN |
| **1997-01-01** | 0.47 | NaN |
| **1998-01-01** | 0.61 | NaN |
| **1999-01-01** | 0.39 | NaN |
| **2000-01-01** | 0.41 | NaN |
| **2001-01-01** | 0.54 | NaN |
| **2002-01-01** | 0.63 | NaN |
| **2003-01-01** | 0.62 | NaN |
| **2004-01-01** | 0.54 | NaN |
| **2005-01-01** | 0.68 | NaN |
| **2006-01-01** | 0.64 | NaN |
| **2007-01-01** | 0.66 | NaN |
| **2008-01-01** | 0.54 | NaN |
| **2009-01-01** | 0.66 | NaN |
| **2010-01-01** | 0.73 | NaN |
| **2011-01-01** | 0.61 | NaN |
| **2012-01-01** | 0.65 | NaN |
| **2013-01-01** | 0.69 | NaN |
| **2014-01-01** | 0.75 | NaN |
| **2015-01-01** | 0.91 | NaN |
| **2016-01-01** | 1.02 | NaN |
| **2017-01-01** | 0.93 | NaN |
| **2018-01-01** | 0.85 | 0.836366 |
| **2019-01-01** | 0.98 | 1.004217 |
| **2020-01-01** | NaN | 0.884745 |
| **2021-01-01** | NaN | 0.842822 |
| **2022-01-01** | NaN | 1.003336 |
| **2023-01-01** | NaN | 1.052633 |
| **2024-01-01** | NaN | 1.085360 |

| | Climate Change( degree celcius) | forecast |
|---|---|---|
| **2025-01-01** | NaN | 1.055382 |
| **2026-01-01** | NaN | 1.076182 |
| **2027-01-01** | NaN | 1.046778 |
| **2028-01-01** | NaN | 1.116956 |
| **2029-01-01** | NaN | 1.182476 |
| **2030-01-01** | NaN | 1.121890 |

In [ ]:
```
1
```

In [ ]:
```
1
```