Pollution MSE

May 31, 2019

```
In [1]: import warnings
        warnings.filterwarnings('ignore')
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        sns.set(font='IPAGothic')
        import numpy as np
        import statsmodels.api as sm
        from sklearn.metrics import mean_squared_error

In [2]: data = pd.read_excel('cleaned_data/Pollution full data Eve.xlsx', parse_dates=['date_tim

In [3]: data = data['pollutantGlobalIndex']

In [4]: print(data.index.min(), data.index.max())

2017-09-06 14:00:00 2019-05-21 12:00:00


In [5]: a = data['2017-09-06 15:00:00':]

In [6]: print(data.index.min(), data.index.max())

2017-09-06 14:00:00 2019-05-21 12:00:00


In [7]: a.head()

Out[7]: date_time
        2017-09-06 15:00:00    3
        2017-09-06 16:00:00    3
        2017-09-06 17:00:00    3
        2017-09-06 18:00:00    3
        2017-09-06 19:00:00    3
        Name: pollutantGlobalIndex, dtype: int64

In [8]: a.count()

Out[8]: 26273
```

```
In [9]: data.count()

Out[9]: 26274

In [10]: data = a

In [11]: data.count()

Out[11]: 26273

In [12]: data[16617:]

Out[12]: date_time
         2018-04-10 01:00:00    5
         2018-04-10 01:00:00    4
         2018-04-10 01:00:00    7
         2018-04-10 02:00:00    7
         2018-04-10 02:00:00    5
         2018-04-10 02:00:00    4
         2018-04-10 02:00:00    7
         2018-04-10 03:00:00    6
         2018-04-10 03:00:00    4
         2018-04-10 03:00:00    4
         2018-04-10 03:00:00    6
         2018-04-10 04:00:00    6
         2018-04-10 04:00:00    4
         2018-04-10 04:00:00    5
         2018-04-10 04:00:00    6
         2018-04-10 05:00:00    6
         2018-04-10 05:00:00    5
         2018-04-10 05:00:00    5
         2018-04-10 05:00:00    6
         2018-04-10 06:00:00    6
         2018-04-10 06:00:00    5
         2018-04-10 06:00:00    5
         2018-04-10 06:00:00    6
         2018-04-10 07:00:00    6
         2018-04-10 07:00:00    5
         2018-04-10 07:00:00    5
         2018-04-10 07:00:00    6
         2018-04-10 09:00:00    6
         2018-04-10 09:00:00    4
         2018-04-10 09:00:00    5
                                ..
         2019-05-20 07:00:00    6
         2019-05-20 08:00:00    6
         2019-05-20 09:00:00    6
         2019-05-20 10:00:00    6
         2019-05-20 11:00:00    6
```

```
2019-05-20 12:00:00    7
2019-05-20 13:00:00    7
2019-05-20 14:00:00    7
2019-05-20 15:00:00    7
2019-05-20 16:00:00    8
2019-05-20 17:00:00    9
2019-05-20 18:00:00    8
2019-05-20 19:00:00    8
2019-05-20 20:00:00    7
2019-05-20 21:00:00    7
2019-05-20 22:00:00    7
2019-05-20 23:00:00    7
2019-05-21 00:00:00    6
2019-05-21 01:00:00    4
2019-05-21 02:00:00    4
2019-05-21 03:00:00    4
2019-05-21 04:00:00    4
2019-05-21 05:00:00    6
2019-05-21 06:00:00    4
2019-05-21 07:00:00    7
2019-05-21 08:00:00    6
2019-05-21 09:00:00    7
2019-05-21 10:00:00    7
2019-05-21 11:00:00    7
2019-05-21 12:00:00    7
Name: pollutantGlobalIndex, Length: 9656, dtype: int64
```

```
In [13]: 33234/2
```

```
Out[13]: 16617.0
```

```
In [14]: data['2019-05-18 15:00:00']
```

```
Out[14]: 6
```

```
In [15]: tr_start, tr_end = '2017-09-06 15:00:00', '2018-03-05 17:00:00'
         te_start, te_end = '2018-03-05 18:00:00', '2019-05-20 11:00:00'
```

```
In [16]: tra = data[tr_start:tr_end].dropna()
         tes = data[te_start:te_end].dropna()
```

```
In [17]: tra.count()
```

```
Out[17]: 13484
```

```
In [18]: tes.count()
```

```
Out[18]: 12764
```

```
In [19]: resDiff = sm.tsa.arma_order_select_ic(tra, max_ar=7, max_ma=7, ic='aic', trend='c')
         print('ARMA(p,q) =',resDiff['aic_min_order'],'is the best.')
```

```
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
```

```
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
```

```
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/base/model.py:508: ConvergenceWarning:
  "Check mle_retvals", ConvergenceWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
```

```
ARMA(p,q) = (6, 7) is the best.
```

Parameters

In [20]:
```
arima = sm.tsa.statespace.SARIMAX(tra,order=(7,1,7),seasonal_order=(0,0,0,0),
                                  enforce_stationarity=False, enforce_invertibility=Fals
arima.summary()
```

```
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:225: ValueWarnin
  ' ignored when e.g. forecasting.', ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/base/model.py:508: ConvergenceWarning:
  "Check mle_retvals", ConvergenceWarning)
```

Out[20]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                              Statespace Model Results
==============================================================================
Dep. Variable:     pollutantGlobalIndex   No. Observations:                13484
Model:                 SARIMAX(7, 1, 7)   Log Likelihood              -16423.938
Date:                  Thu, 30 May 2019   AIC                          32877.875
Time:                          23:51:48   BIC                          32990.504
Sample:                               0   HQIC                         32915.441
                                - 13484
Covariance Type:                    opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.7667      0.059    -29.753      0.000      -1.883      -1.650
ar.L2         -1.3923      0.084    -16.635      0.000      -1.556      -1.228
ar.L3         -0.6524      0.068     -9.619      0.000      -0.785      -0.519
ar.L4          0.6252      0.033     18.940      0.000       0.561       0.690
ar.L5          1.3118      0.045     28.884      0.000       1.223       1.401
ar.L6          0.9910      0.054     18.214      0.000       0.884       1.098
ar.L7          0.3332      0.039      8.553      0.000       0.257       0.410
ma.L1          1.0844      0.057     18.891      0.000       0.972       1.197
ma.L2         -0.1875      0.059     -3.166      0.002      -0.304      -0.071
ma.L3         -0.4530      0.049     -9.265      0.000      -0.549      -0.357
ma.L4         -0.4643      0.041    -11.279      0.000      -0.545      -0.384
ma.L5         -0.4412      0.026    -17.145      0.000      -0.492      -0.391
ma.L6         -0.2515      0.031     -8.120      0.000      -0.312      -0.191
ma.L7         -0.0843      0.031     -2.751      0.006      -0.144      -0.024
sigma2         0.6528      0.007     98.079      0.000       0.640       0.666
==============================================================================
Ljung-Box (Q):                       85.82   Jarque-Bera (JB):             13107.39
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.92   Skew:                             0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.81
==============================================================================

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

pred = arima.predict(tr_end,te_end, dynamic=True)[1:] print('ARIMA model MSE:{}'.format(mean_squared_error(tes,pred)))

```
In [28]: pred = arima.predict(20470,33234, dynamic=True)[1:]
         print('ARIMA model MSE:{}'.format(mean_squared_error(tes,pred)))
```

```
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/base/tsa_model.py:531: ValueWarnin
  ValueWarning)
/home/omar/.local/lib/python3.5/site-packages/statsmodels/tsa/statespace/kalman_filter.py:1740:
  ' effect.', ValueWarning)
```

```
ARIMA model MSE:4.0156838160913315
```