Untitled1

July 6, 2020

[1]:
```
### FITTING MULTI LINEAR REGRESSION MODEL FOR MUCROINSURANCE DATASET
```

[2]:
```
## Modules required
import pandas as pd
import seaborn as sns
import numpy as np
import pylab
import math
import matplotlib.pyplot as plt
```

[3]:
```
from scipy import stats
import statsmodels.api as sm
from statsmodels.stats import diagnostic as diag
from statsmodels.stats.outliers_influence import variance_inflation_factor

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error

%matplotlib inline
```

[4]:
```
## Load the dataset into pandas
MICDATA=pd.read_excel('MIP.xlsx')

## set the index equal to the year column
MICDATA.index = MICDATA['Year']
MICDATA = MICDATA.drop('Year', axis = 1)

MICDATA.head()
```

[5]:
```
## Checking the Model Assumptions
######## Multicolinearity ################
## printing out correlation matrix of the data frame
corr=MICDATA.corr()

## Display the correlation matrix
display(corr)
```

```python
## plot a heatmap
sns.heatmap(corr, xticklabels = corr.columns, yticklabels = corr.columns, cmap="RdBu")
```

```
               ADR     PSIZE       PGR     PPOOR     PPLRA      MCTS  \
ADR       1.000000  0.030680 -0.611861 -0.556588 -0.616613  0.498074
PSIZE     0.030680  1.000000  0.094497  0.065800  0.024948 -0.138829
PGR      -0.611861  0.094497  1.000000  0.325237  0.696683 -0.709433
PPOOR    -0.556588  0.065800  0.325237  1.000000  0.447905 -0.435525
PPLRA    -0.616613  0.024948  0.696683  0.447905  1.000000 -0.875357
MCTS      0.498074 -0.138829 -0.709433 -0.435525 -0.875357  1.000000
PIUI      0.695094  0.124953 -0.678652 -0.492520 -0.788620  0.790480
EGROWTH   0.519087  0.001746 -0.544261 -0.404348 -0.698102  0.709602
 LEB      0.719510 -0.184154 -0.493371 -0.648189 -0.300559  0.344735
EODB      0.030722 -0.079051 -0.032027  0.000098  0.002121  0.058295
GNIPC    -0.018822 -0.131520  0.004033  0.040234  0.011001  0.045133
INFLAT   -0.106615 -0.125535 -0.035370  0.168146  0.034730  0.053770
REALIR   -0.013517  0.148069  0.098491 -0.062883  0.133606 -0.198379
GDPOMT    0.027404 -0.197347 -0.039898  0.024243 -0.081999  0.125092
INSP      0.114073 -0.141601 -0.124487 -0.049185 -0.116061  0.152582
INSDENS   0.094796 -0.126445 -0.086087 -0.053918 -0.084251  0.124674
BUSFRD   -0.042811 -0.093083  0.056206  0.042807  0.048292  0.057419
 CPI      0.260820 -0.333092 -0.309719 -0.433986 -0.436487  0.570107
OPOE      0.020813  0.038039  0.067650 -0.046260  0.085067 -0.059058
FISFRED  -0.022139  0.083895  0.086104  0.009585  0.051734 -0.029266
FINSF    -0.019923  0.067335  0.078981 -0.000651  0.031461 -0.008479
PROPR     0.093058  0.066310 -0.026464 -0.079693 -0.045826  0.039277
GOVSPEND  0.108640 -0.092637 -0.072985 -0.046747 -0.079764  0.073630

              PIUI   EGROWTH       LEB      EODB  ...    GDPOMT      INSP  \
ADR       0.695094  0.519087  0.719510  0.030722  ...  0.027404  0.114073
PSIZE     0.124953  0.001746 -0.184154 -0.079051  ... -0.197347 -0.141601
PGR      -0.678652 -0.544261 -0.493371 -0.032027  ... -0.039898 -0.124487
PPOOR    -0.492520 -0.404348 -0.648189  0.000098  ...  0.024243 -0.049185
PPLRA    -0.788620 -0.698102 -0.300559  0.002121  ... -0.081999 -0.116061
MCTS      0.790480  0.709602  0.344735  0.058295  ...  0.125092  0.152582
PIUI      1.000000  0.601579  0.523728  0.063156  ...  0.105522  0.187925
EGROWTH   0.601579  1.000000  0.363433 -0.012884  ...  0.094272  0.037928
 LEB      0.523728  0.363433  1.000000  0.073294  ...  0.045361  0.114229
EODB      0.063156 -0.012884  0.073294  1.000000  ...  0.251357  0.351837
GNIPC    -0.042500 -0.078773 -0.036030  0.479541  ...  0.333786  0.531417
INFLAT   -0.009023 -0.051479 -0.051917 -0.017658  ... -0.018734 -0.033881
REALIR   -0.127980 -0.027585  0.000287  0.043872  ... -0.200363 -0.152972
GDPOMT    0.105522  0.094272  0.045361  0.251357  ...  1.000000  0.370100
INSP      0.187925  0.037928  0.114229  0.351837  ...  0.370100  1.000000
INSDENS   0.143838  0.011025  0.098219  0.358797  ...  0.301863  0.966102
BUSFRD    0.001449  0.018159 -0.012376  0.509050  ...  0.386968  0.330940
```

CPI	0.371850	0.505284	0.367819	0.061592	...	0.055125	0.131155
OPOE	-0.062065	-0.053190	0.021048	0.529251	...	0.012781	0.186543
FISFRED	-0.139834	0.010781	-0.084135	0.010346	...	-0.499965	-0.465369
FINSF	-0.141906	0.027759	-0.082181	0.007598	...	-0.504960	-0.467978
PROPR	0.113285	-0.010861	0.037888	0.526250	...	0.180632	0.344774
GOVSPEND	0.085034	-0.007365	0.096475	0.220083	...	-0.136880	0.664856

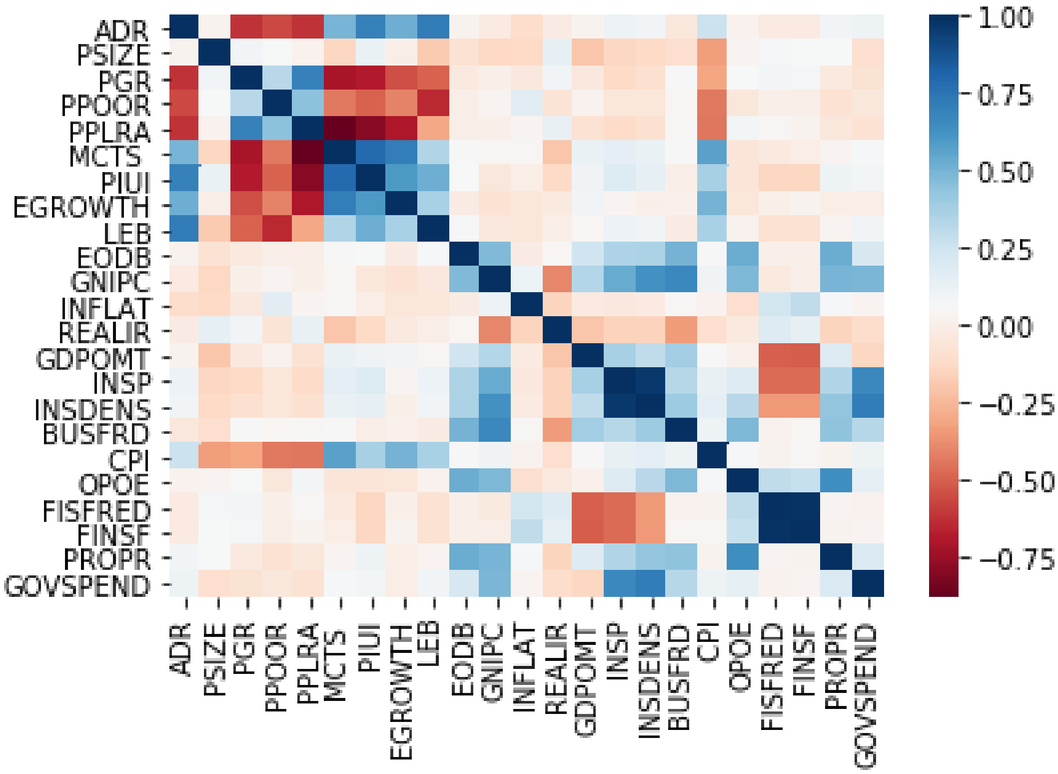
	INSDENS	BUSFRD	CPI	OPOE	FISFRED	FINSF \
ADR	0.094796	-0.042811	0.260820	0.020813	-0.022139	-0.019923
PSIZE	-0.126445	-0.093083	-0.333092	0.038039	0.083895	0.067335
PGR	-0.086087	0.056206	-0.309719	0.067650	0.086104	0.078981
PPOOR	-0.053918	0.042807	-0.433986	-0.046260	0.009585	-0.000651
PPLRA	-0.084251	0.048292	-0.436487	0.085067	0.051734	0.031461
MCTS	0.124674	0.057419	0.570107	-0.059058	-0.029266	-0.008479
PIUI	0.143838	0.001449	0.371850	-0.062065	-0.139834	-0.141906
EGROWTH	0.011025	0.018159	0.505284	-0.053190	0.010781	0.027759
LEB	0.098219	-0.012376	0.367819	0.021048	-0.084135	-0.082181
EODB	0.358797	0.509050	0.061592	0.529251	0.010346	0.007598
GNIPC	0.628384	0.669485	0.103230	0.481145	-0.030986	-0.000523
INFLAT	-0.023006	0.050831	0.020106	-0.096189	0.245388	0.297404
REALIR	-0.155562	-0.335807	-0.097743	-0.028216	0.180800	0.147824
GDPOMT	0.301863	0.386968	0.055125	0.012781	-0.499965	-0.504960
INSP	0.966102	0.330940	0.131155	0.186543	-0.465369	-0.467978
INSDENS	1.000000	0.402478	0.150277	0.322487	-0.342165	-0.342629
BUSFRD	0.402478	1.000000	0.108302	0.484022	0.023516	0.046795
CPI	0.150277	0.108302	1.000000	0.072600	0.029044	0.053964
OPOE	0.322487	0.484022	0.072600	1.000000	0.297488	0.276972
FISFRED	-0.342165	0.023516	0.029044	0.297488	1.000000	0.987854
FINSF	-0.342629	0.046795	0.053964	0.276972	0.987854	1.000000
PROPR	0.429160	0.438530	0.030881	0.643603	0.033641	0.025839
GOVSPEND	0.719676	0.327525	0.110510	0.145274	0.018759	0.034729

	PROPR	GOVSPEND
ADR	0.093058	0.108640
PSIZE	0.066310	-0.092637
PGR	-0.026464	-0.072985
PPOOR	-0.079693	-0.046747
PPLRA	-0.045826	-0.079764
MCTS	0.039277	0.073630
PIUI	0.113285	0.085034
EGROWTH	-0.010861	-0.007365
LEB	0.037888	0.096475
EODB	0.526250	0.220083
GNIPC	0.497330	0.489741
INFLAT	0.074500	0.036190
REALIR	-0.144625	-0.112977
GDPOMT	0.180632	-0.136880
INSP	0.344774	0.664856

```
INSDENS   0.429160  0.719676
BUSFRD    0.438530  0.327525
 CPI      0.030881  0.110510
OPOE      0.643603  0.145274
FISFRED   0.033641  0.018759
FINSF     0.025839  0.034729
PROPR     1.000000  0.195215
GOVSPEND  0.195215  1.000000

[23 rows x 23 columns]
```

```
[5]: <matplotlib.axes._subplots.AxesSubplot at 0x19e127bebc8>
```




```
[7]: ### Using the VIF to measure to detect the above
MICDATA_before = MICDATA
#MICDATA_after = MICDATA.drop(['PPLRA','PIUI','FINSF'], axis = 1)

x1 = sm.tools.add_constant(MICDATA_before)
#x2 = sm.tools.add_constant(MICDATA_after)

# Create a series for both
```

```
series_before = pd.Series([variance_inflation_factor(x1.values, i) for i in␣
 ↪range(x1.shape[1])], index = x1.columns)
#series_after = pd.Series([variance_inflation_factor(x2.values, i) for i in␣
 ↪range(x2.shape[1])], index = x2.columns)

## dispay the series
print('DATA BEFORE')
print('-'*100)
display(series_before)

#print('DATA AFTER')
#print('-'*100)
#display(series_after)
```

```
DATA BEFORE
--------------------------------------------------------------------------------
--------------------

const       1103.067273
ADR            4.542684
PSIZE          1.759610
PGR            3.126149
PPOOR          2.528102
PPLRA          8.320064
MCTS           9.537138
PIUI           6.561252
EGROWTH        2.654841
 LEB           4.740164
EODB           2.542639
GNIPC          4.227378
INFLAT         1.786472
REALIR         1.805792
GDPOMT         3.039639
INSP          35.926502
INSDENS       41.304398
BUSFRD         2.931325
 CPI           2.142790
OPOE           3.638434
FISFRED       61.474058
FINSF         70.118553
PROPR          2.436500
GOVSPEND       4.858222
dtype: float64
```

[ ]:
```
## Visualizing the DATA AFTER on scatter plot
pd.plotting.scatter_matrix(MICDATA_after, alpha = 1, figsize = (30, 20))
```

```
## show the plot
plt.show()
```

[ ]:
```
## Get the summary of our original data set
desc_MICDATA = MICDATA.describe()

## Add the standard deviation metric
desc_MICDATA.loc['+3_std']=desc_MICDATA.loc['mean']+(desc_MICDATA.loc['std']*3)
desc_MICDATA.loc['-3_std']=desc_MICDATA.loc['mean']-(desc_MICDATA.loc['std']*3)

desc_MICDATA
```

[ ]:
```
#### Building the model ####
## considering Economic growth as our dependent Variable ##
## MODEL 5
## define our input variable and our output variable where ###
x = MICDATA_after.drop(['EGROWTH','EODB'], axis = 1)

y= MICDATA_after
```

[ ]:
```
```

[ ]:
```
```