Ela_test1-Copy1

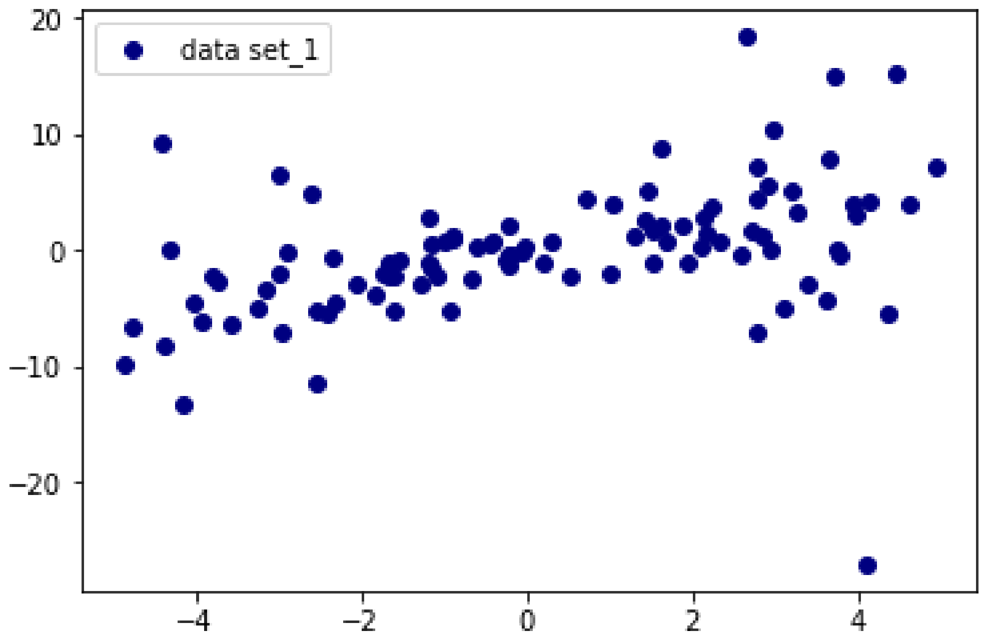
March 11, 2019

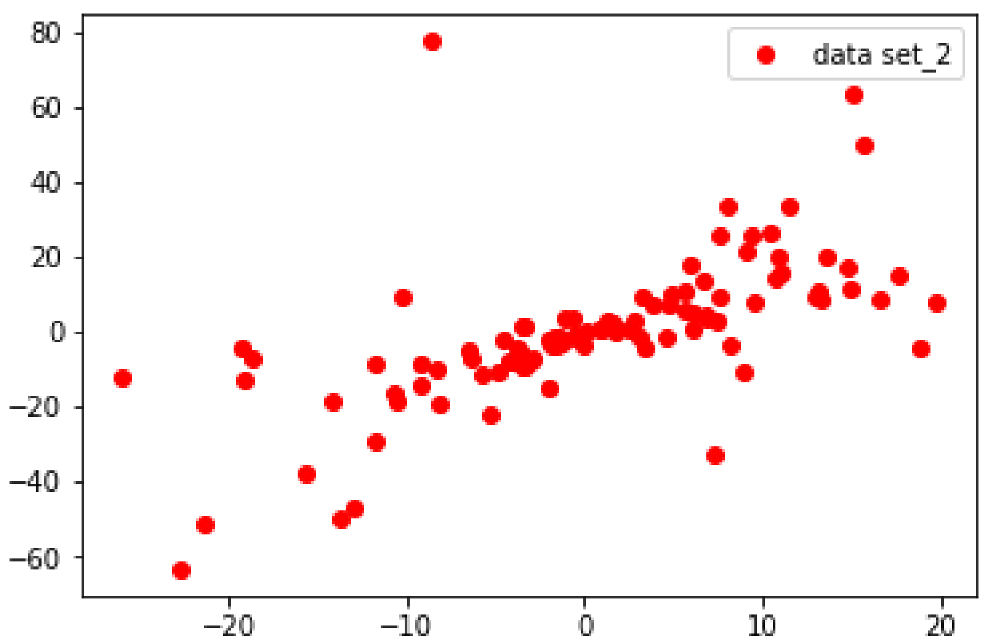
```
In [1]: %matplotlib inline

In [2]: #attaching packages
        from matplotlib import pyplot as plt
        import numpy as np
        import pandas as pd
        import pymc3 as pm
        from scipy import stats
        import seaborn as sns
        import statsmodels
        from statsmodels import api as sm
        import statsmodels.stats.api as sms
        import math
        import statsmodels.api as smv
```

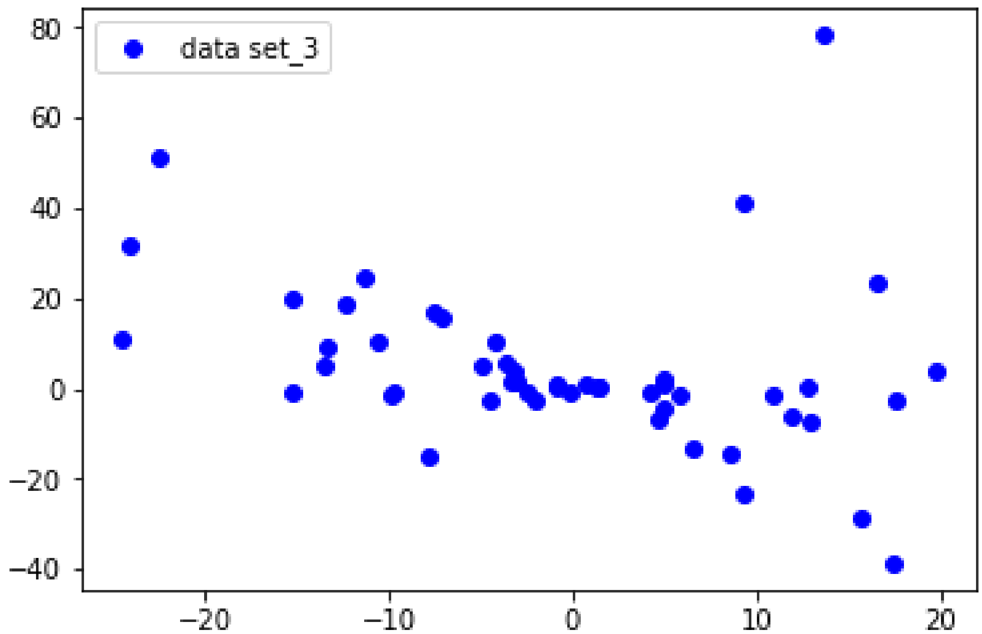
```
WARNING (theano.configdefaults): g++ not available, if using conda: `conda install m2w64-toolcha
C:\Users\Ela\Anaconda3\lib\site-packages\theano\configdefaults.py:560: UserWarning: DeprecationW
  warnings.warn("DeprecationWarning: there is no c++ compiler."
WARNING (theano.configdefaults): g++ not detected ! Theano will be unable to execute optimized C
WARNING (theano.tensor.blas): Using NumPy C-API based implementation for BLAS functions.
```

```
In [3]: #loading 5 data sets from home-directory
        data = {}
        resp = {}
        for i in range(1,6):
            fname='X_'+str(i)+'.txt'
            fnamey='Y_'+str(i)+'.txt'
            X=np.loadtxt(fname)
            Y=np.loadtxt(fnamey)
            data[i-1] = X
            resp[i-1] = Y

        MINN=np.zeros(5)
        MAXX=np.zeros(5)
        C=['navy', 'red', 'blue', 'green', 'purple','yellow']
        for i in range(0,5):
            MINN[i], MAXX[i] = data[i].min() - 1, data[i].max() + 1
```

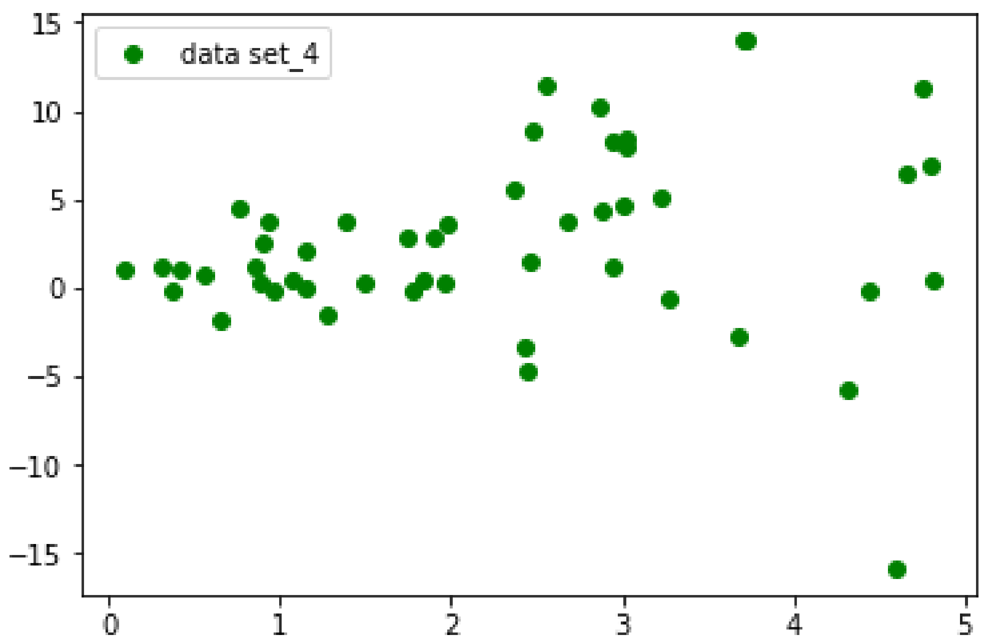
```
fig, ax = plt.subplots()
ax.scatter(data[i], resp[i],color=C[i]);
ax.legend(['data set_'+str(i+1)])
```



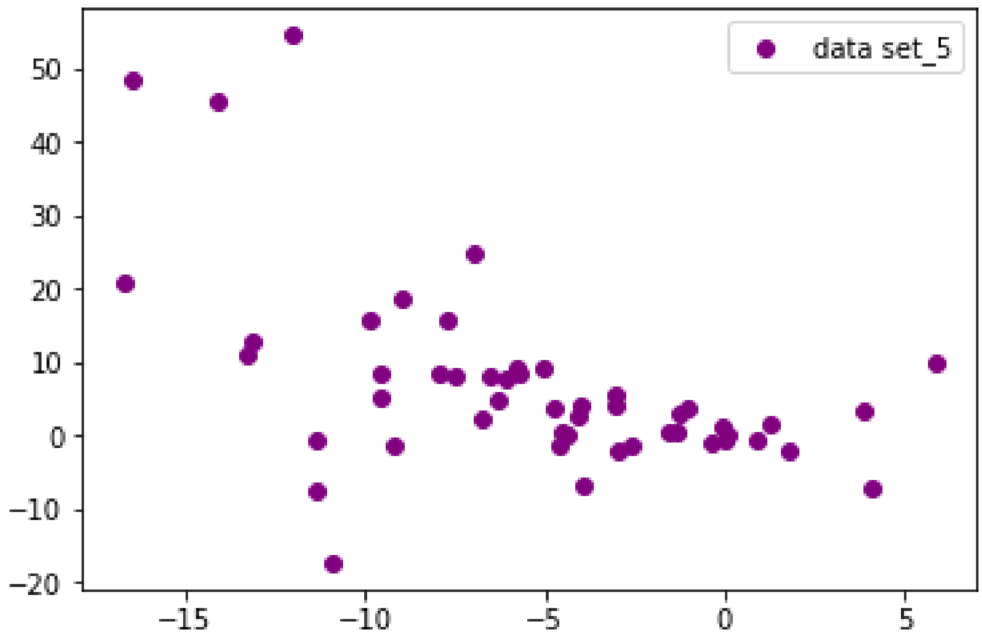




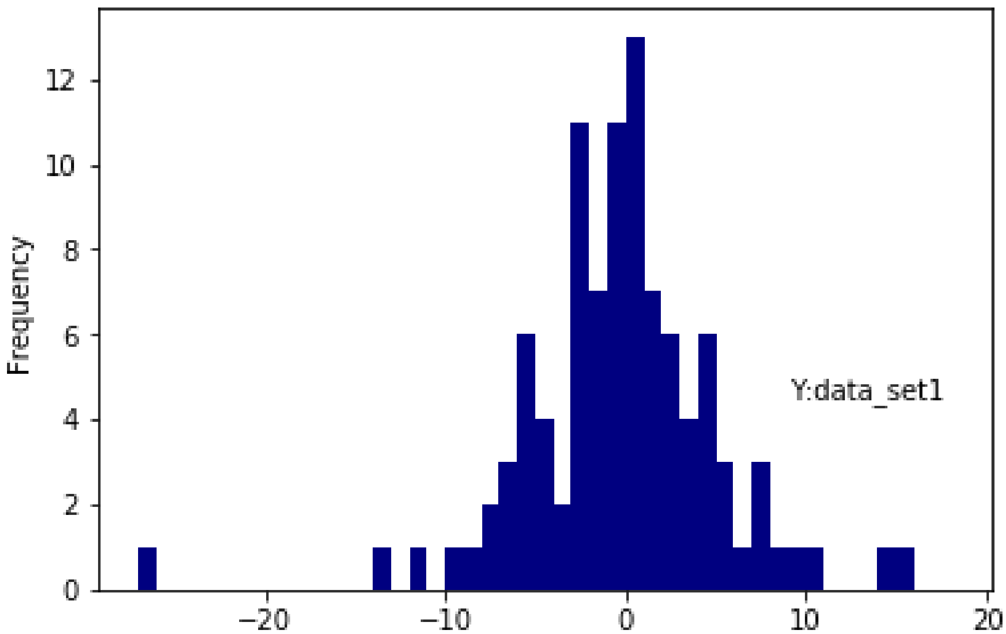
data set_3
80
60
40
20
0
−20
−40
−20
−10
0
10
20

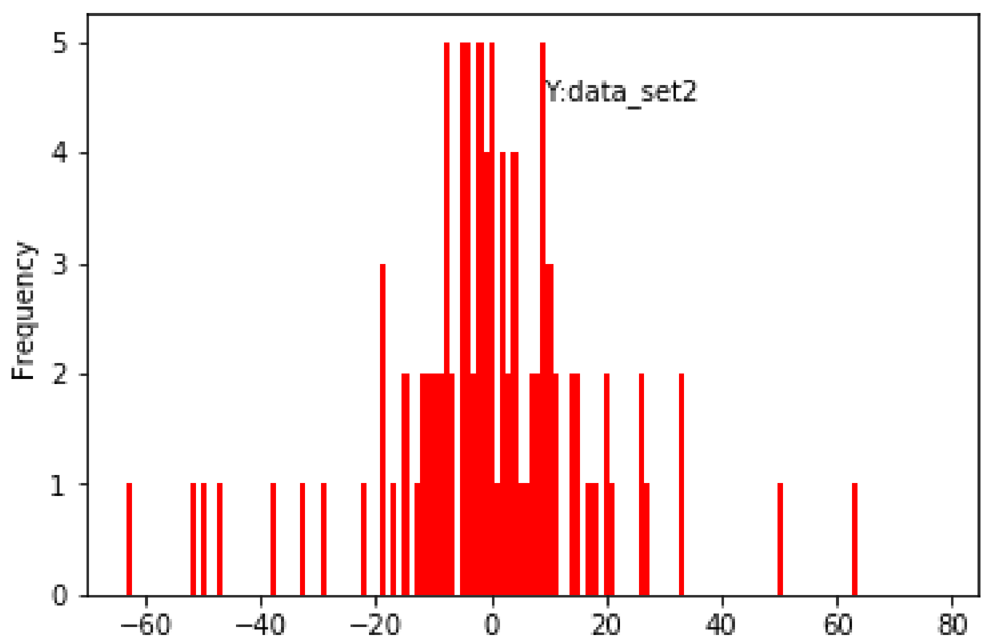


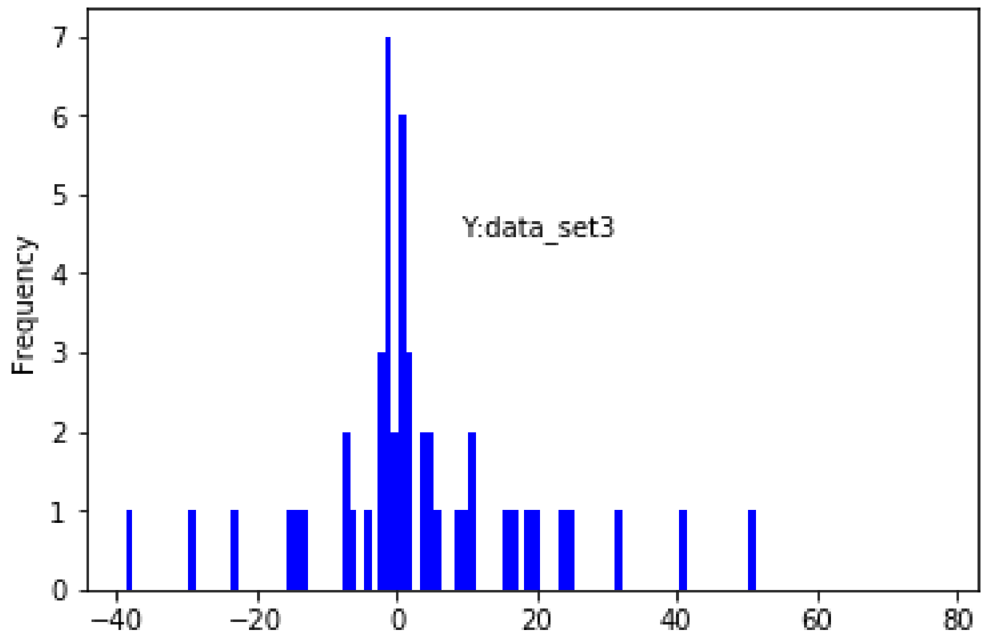
data set_4
15
10
5
0
−5
−10
−15
0
1
2
3
4
5



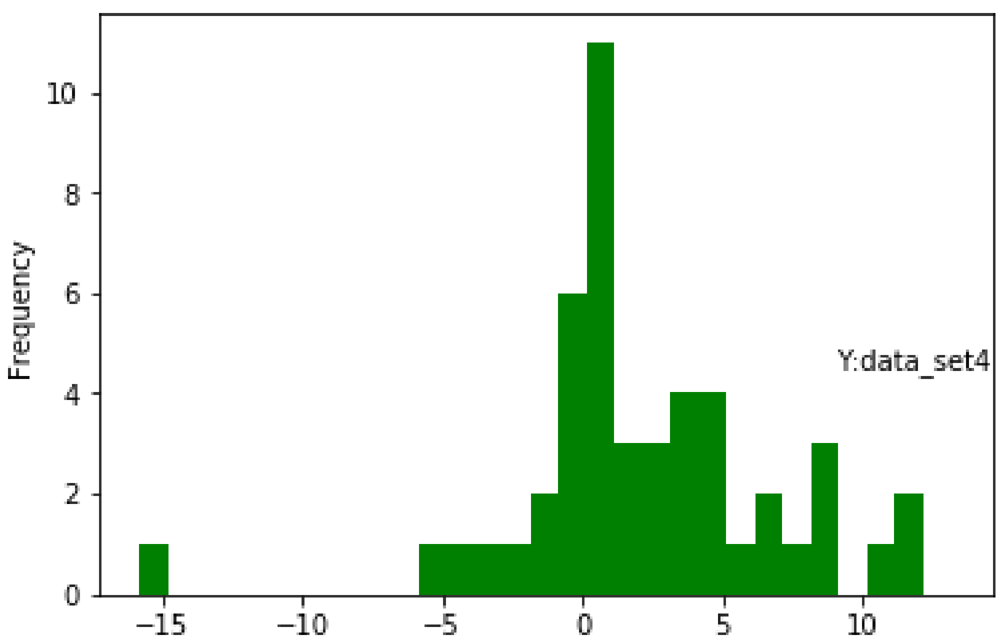

```
In [4]: #distribution of Y
        MINNR=np.zeros(5)
        MAXXR=np.zeros(5)
        for i in range(0,5):
            MINNR[i], MAXXR[i] =resp[i].min(), resp[i].max()
            BINNR= list(np.arange(MINNR[i],MAXXR[i],1))
            fig, ax = plt.subplots()
            pd.Series(resp[i]).plot(kind='hist', bins=BINNR,color=C[i],label="KK")
            plt.text(9, 4.5, 'Y:data_set'+str(i+1))
```


Frequency
Y:data_set1
0
2
4
6
8
10
12
−20
−10
0
10
20

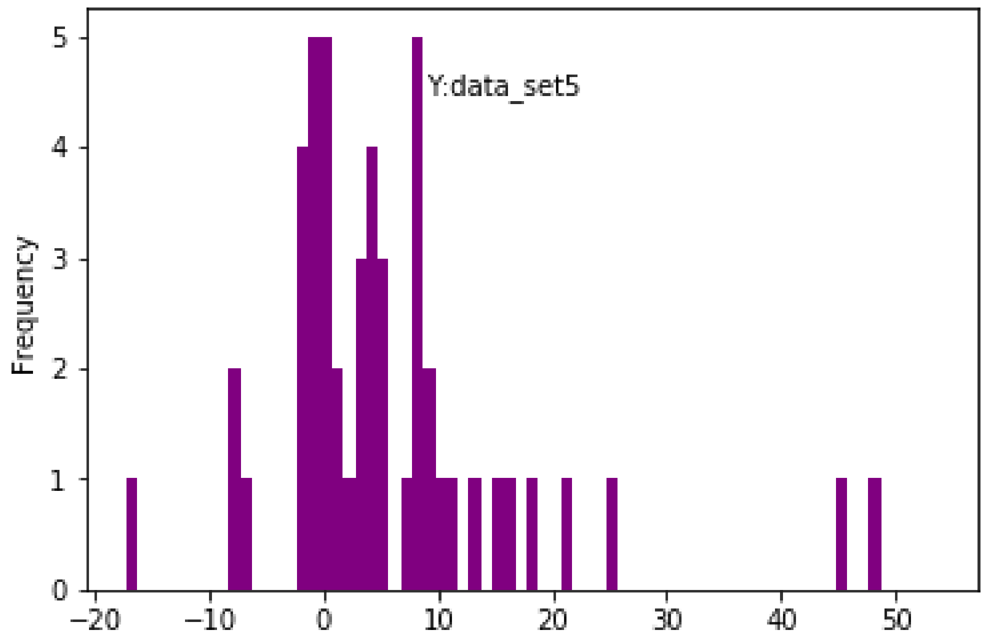

Frequency
Y:data_set2
0
1
2
3
4
5
−60
−40
−20
0
20
40
60
80



Y:data_set3
Frequency
0
1
2
3
4
5
6
7
−40
−20
0
20
40
60
80



Y:data_set4
Frequency
0
2
4
6
8
10
−15
−10
−5
0
5
10




```
In [5]: #OLS model
        dataP={}
        ols_result={}
        for i in range(0,5):
            dataP[i]=sm.add_constant(data[i])
            #print(dataP[i])
            ols_result[i] = sm.OLS(resp[i],dataP[i] ).fit()
            print(ols_result[i].summary())
            #print('Parameters: ', ols_result[i].params)
            #print('R2: ', ols_result[i].rsquared)
```

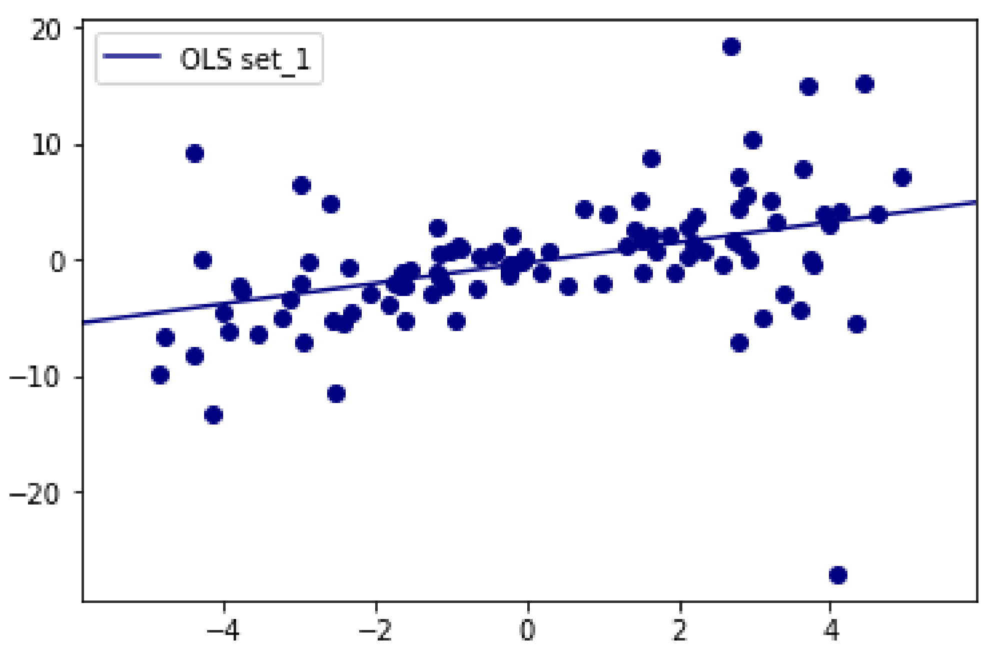
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.164
Model:                            OLS   Adj. R-squared:                  0.155
Method:                 Least Squares   F-statistic:                     19.22
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           2.93e-05
Time:                        20:16:19   Log-Likelihood:                -308.57
No. Observations:                 100   AIC:                             621.1
Df Residuals:                      98   BIC:                             626.3
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.2695      0.536     -0.503      0.616      -1.333       0.794
```

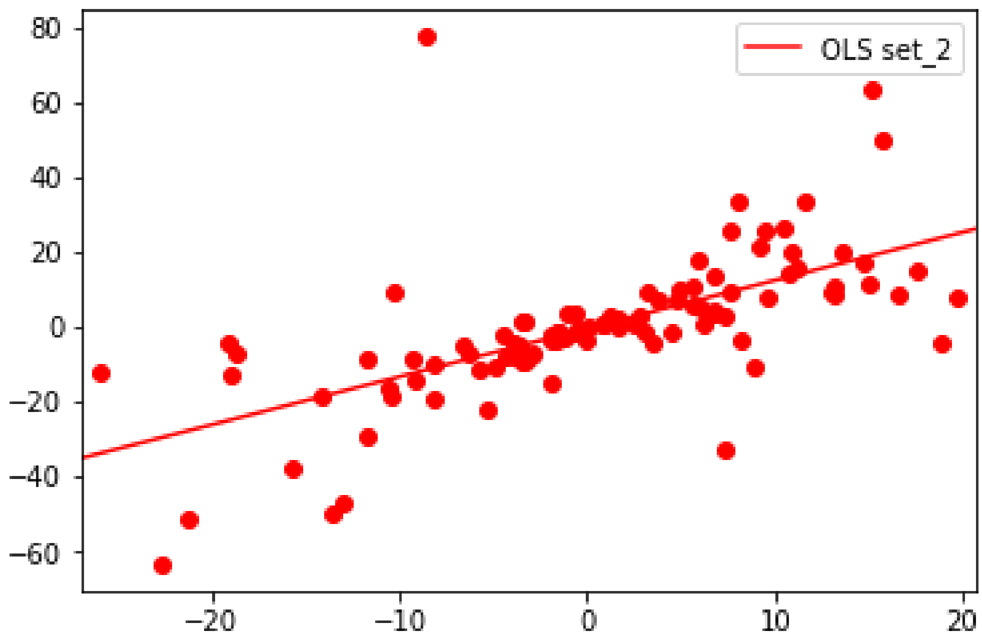
```
x1             0.8808      0.201      4.384      0.000       0.482       1.279
==============================================================================
Omnibus:                       51.887   Durbin-Watson:                   2.329
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              472.400
Skew:                          -1.324   Prob(JB):                    2.63e-103
Kurtosis:                      13.313   Cond. No.                         2.68
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.398
Model:                            OLS   Adj. R-squared:                  0.392
Method:                 Least Squares   F-statistic:                     64.81
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           1.98e-12
Time:                        20:16:19   Log-Likelihood:                -415.48
No. Observations:                 100   AIC:                             835.0
Df Residuals:                      98   BIC:                             840.2
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.3959      1.560     -0.254      0.800      -3.491       2.700
x1             1.2856      0.160      8.051      0.000       0.969       1.602
==============================================================================
Omnibus:                       62.632   Durbin-Watson:                   2.181
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              522.315
Skew:                           1.778   Prob(JB):                    3.81e-114
Kurtosis:                      13.616   Cond. No.                         9.78
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.103
Model:                            OLS   Adj. R-squared:                  0.085
Method:                 Least Squares   F-statistic:                     5.524
Date:                Sun, 10 Mar 2019   Prob (F-statistic):             0.0229
Time:                        20:16:19   Log-Likelihood:                -214.11
No. Observations:                  50   AIC:                             432.2
Df Residuals:                      48   BIC:                             436.0
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
```

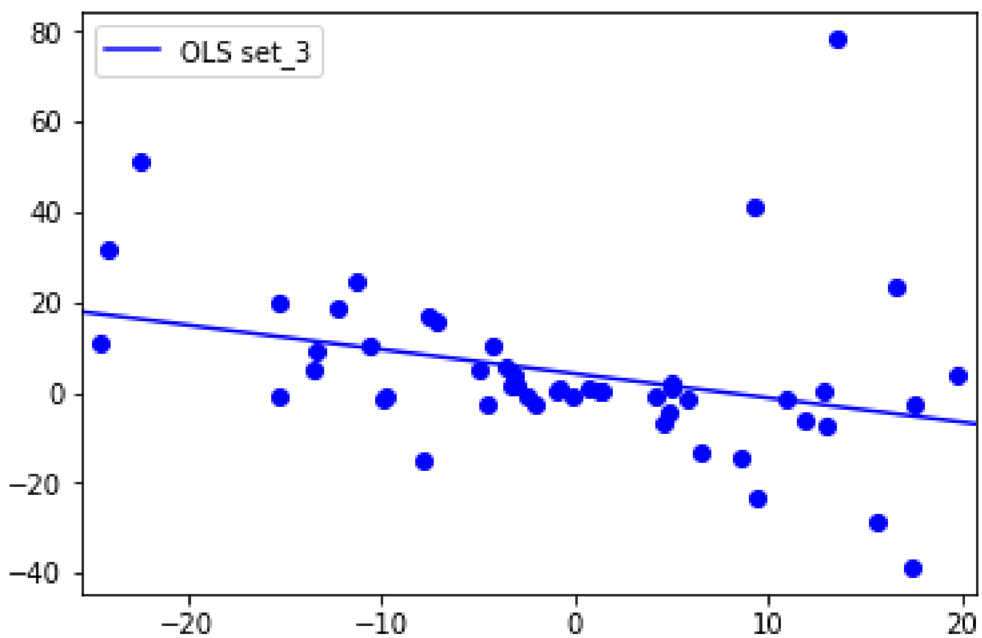
```
==============================================================================
const          4.1752      2.530      1.650      0.105      -0.912       9.263
x1            -0.5378      0.229     -2.350      0.023      -0.998      -0.078
==============================================================================
Omnibus:                       43.071   Durbin-Watson:                   1.922
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              173.506
Skew:                           2.225   Prob(JB):                     2.11e-38
Kurtosis:                      10.968   Cond. No.                         11.1
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.028
Model:                            OLS   Adj. R-squared:                  0.007
Method:                 Least Squares   F-statistic:                     1.362
Date:                Sun, 10 Mar 2019   Prob (F-statistic):              0.249
Time:                        20:16:19   Log-Likelihood:                -152.15
No. Observations:                  50   AIC:                             308.3
Df Residuals:                      48   BIC:                             312.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.1481      1.443      0.796      0.430      -1.753       4.049
x1             0.6334      0.543      1.167      0.249      -0.458       1.725
==============================================================================
Omnibus:                       15.626   Durbin-Watson:                   1.936
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               29.405
Skew:                          -0.848   Prob(JB):                     4.12e-07
Kurtosis:                       6.352   Cond. No.                         5.81
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.281
Model:                            OLS   Adj. R-squared:                  0.266
Method:                 Least Squares   F-statistic:                     18.72
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           7.63e-05
Time:                        20:16:19   Log-Likelihood:                -191.30
No. Observations:                  50   AIC:                             386.6
Df Residuals:                      48   BIC:                             390.4
Df Model:                           1
Covariance Type:            nonrobust
```

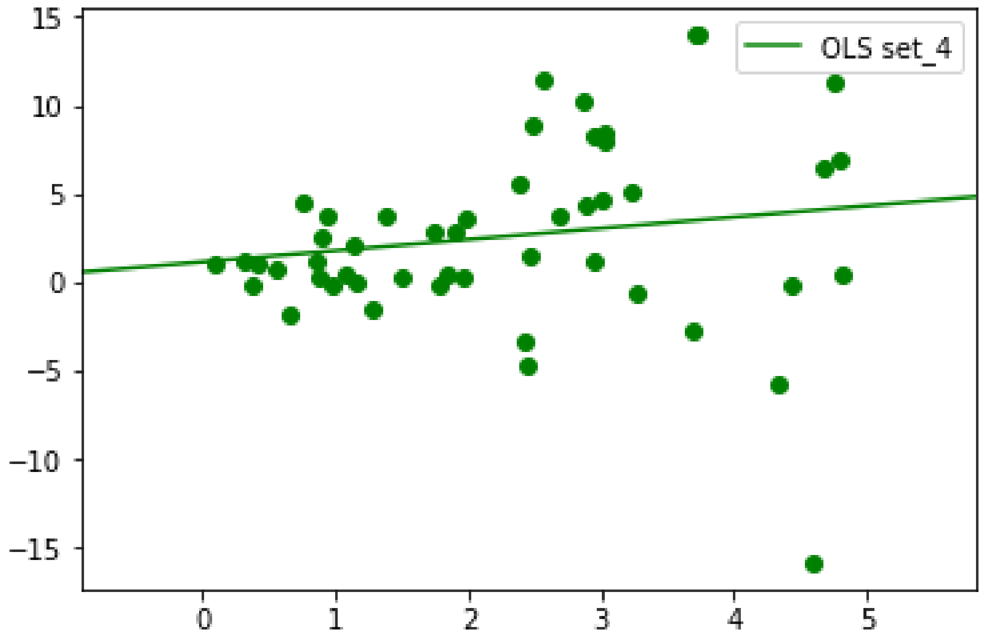
```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.2413      2.302     -0.105      0.917      -4.870       4.387
x1            -1.3331      0.308     -4.327      0.000      -1.953      -0.714
==============================================================================
Omnibus:                       15.549   Durbin-Watson:                   2.048
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               29.921
Skew:                           0.830   Prob(JB):                     3.18e-07
Kurtosis:                       6.407   Cond. No.                         10.8
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

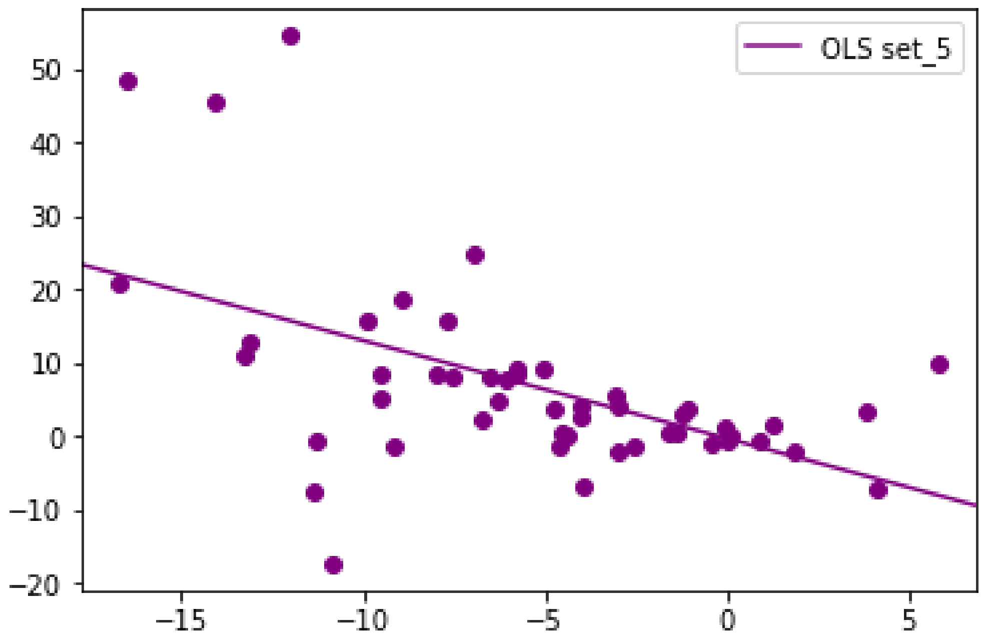
In [6]:
```python
datas={}
Datas={}
for i in range(0,5):
    datas[i]= np.linspace(MINN[i], MAXX[i], 20)
    Datas[i]= sm.add_constant(datas[i])
    fig, ax = plt.subplots()
    ax.scatter(data[i], resp[i],color=C[i]);
    ax.plot(datas[i], ols_result[i].predict(Datas[i]), color=C[i],label='OLS model');
    ax.set_xlim(MINN[i], MAXX[i]);
    ax.legend(loc='upper right');
    ax.legend(['OLS set_'+str(i+1)])
```




OLS set_2
80
60
40
20
0
−20
−40
−60
−20
−10
0
10
20


OLS set_3
80
60
40
20
0
−20
−40
−20
−10
0
10
20





In [7]:
```
#test for normality
print("Test For Checking Normality BY Jarque-Bera")
for i in range(0,5):
```

```
        name = ['Jarque-Bera', 'Chi^2 two-tail prob.', 'Skew', 'Kurtosis']
        test = sms.jarque_bera(ols_result[i].resid)
        print('data_'+str(i+1), test)
```

```
Test For Checking Normality BY Jarque-Bera
data_1 (472.39980463722924, 2.6283657356926387e-103, -1.324067850511277, 13.313259931850103)
data_2 (522.3149203148531, 3.808519168472229e-114, 1.778461703504943, 13.61620820972488)
data_3 (173.5060597283974, 2.1068704967045583e-38, 2.224833573047068, 10.96764525807978)
data_4 (29.40467998416821, 4.119598301096693e-07, -0.8479323772682648, 6.352355757780116)
data_5 (29.921077427200533, 3.1821495811975905e-07, 0.8301034559594401, 6.40673306478113)
```

```
In [8]: print('Heteroscedasticity Test')
        ols_resid={}
        resid_fit={}
        rho={}
        for i in range(0,5):
            ols_resid[i] = ols_result[i].resid
            resid_fit[i] = sm.OLS(ols_resid[i][1:], sm.add_constant(ols_resid[i][:-1])).fit()
            rho[i] = resid_fit[i].params[1]

            #test for Heteroscedasticity Test
        for i in range(0,5):
            print('data_set_'+str(i+1),statsmodels.stats.diagnostic.het_white(ols_resid[i]**2, d
```

```
Heteroscedasticity Test
data_set_1 (4.808483341748426, 0.09033397252489685, 2.449918335811942, 0.0916252027127786)
data_set_2 (0.8001970719581641, 0.6702539986476804, 0.39122616017818734, 0.677289362845173)
data_set_3 (2.601350287860371, 0.2723478569468163, 1.289735723190059, 0.28490835576732)
data_set_4 (6.019181004685592, 0.04931186772081809, 3.216191895498382, 0.04907866478928356)
data_set_5 (7.6502982613980475, 0.021815181975108996, 4.245177693399951, 0.020192415053641378)
```

```
In [12]: print('Outliers Identification')
         ###using Huber's T norm with the default

         huber_t={}
         hub_results={}
         for i in range(0,5):
             huber_t [i]= smv.RLM(resp[i], dataP[i], M=smv.robust.norms.HuberT())
             hub_results[i] = huber_t[i].fit()
             print('min weight for data_set_'+str(i+1),hub_results[i].weights.min())
```

```
Outliers Identification
min weight for data_set_1 0.12089950165960449
min weight for data_set_2 0.10322183897415442
min weight for data_set_3 0.13476214428731814
```

```
min weight for data_set_4 0.19749328013764936
min weight for data_set_5 0.14218929521869575
```

In [13]:
```
#GLS modeling
from scipy.linalg import toeplitz
order={}
sigma={}
gls_model={}
gls_results={}

toeplitz(range(5))
for i in range(0,5):
    order[i] = toeplitz(range(len(ols_resid[i])))
    sigma[i] = rho[i]**order[i]
    gls_model[i] = sm.GLS(resp[i],dataP[i], sigma=sigma[i])
    gls_results[i] = gls_model[i].fit()
    print(gls_results[i].summary())
```

```
                            GLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.185
Model:                            GLS   Adj. R-squared:                  0.176
Method:                 Least Squares   F-statistic:                     22.19
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           8.13e-06
Time:                        20:20:42   Log-Likelihood:                -307.14
No. Observations:                 100   AIC:                             618.3
Df Residuals:                      98   BIC:                             623.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.2836      0.454     -0.625      0.533      -1.184       0.616
x1             0.9179      0.195      4.710      0.000       0.531       1.305
==============================================================================
Omnibus:                       45.336   Durbin-Watson:                   2.010
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              351.303
Skew:                          -1.148   Prob(JB):                     5.20e-77
Kurtosis:                      11.891   Cond. No.                         2.34
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            GLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.420
```

```
Model:                            GLS   Adj. R-squared:                  0.414
Method:                 Least Squares   F-statistic:                     70.87
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           3.22e-13
Time:                        20:20:42   Log-Likelihood:                -414.99
No. Observations:                 100   AIC:                             834.0
Df Residuals:                      98   BIC:                             839.2
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.4317      1.417     -0.305      0.761      -3.244       2.381
x1             1.3242      0.157      8.418      0.000       1.012       1.636
==============================================================================
Omnibus:                       63.699   Durbin-Watson:                   2.002
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              536.745
Skew:                           1.816   Prob(JB):                    2.80e-117
Kurtosis:                      13.753   Cond. No.                         9.02
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            GLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.103
Model:                            GLS   Adj. R-squared:                  0.084
Method:                 Least Squares   F-statistic:                     5.514
Date:                Sun, 10 Mar 2019   Prob (F-statistic):             0.0230
Time:                        20:20:42   Log-Likelihood:                -214.08
No. Observations:                  50   AIC:                             432.2
Df Residuals:                      48   BIC:                             436.0
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.1835      2.622      1.595      0.117      -1.089       9.456
x1            -0.5388      0.229     -2.349      0.023      -1.000      -0.078
==============================================================================
Omnibus:                       43.252   Durbin-Watson:                   1.984
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              173.759
Skew:                           2.240   Prob(JB):                     1.86e-38
Kurtosis:                      10.959   Cond. No.                         11.4
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            GLS Regression Results
```

```
==============================================================================
Dep. Variable:                      y   R-squared:                       0.025
Model:                            GLS   Adj. R-squared:                  0.005
Method:                 Least Squares   F-statistic:                     1.243
Date:                Sun, 10 Mar 2019   Prob (F-statistic):              0.270
Time:                        20:20:42   Log-Likelihood:                -152.12
No. Observations:                  50   AIC:                             308.2
Df Residuals:                      48   BIC:                             312.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.2190      1.454      0.839      0.406      -1.704       4.142
x1             0.6028      0.543      1.110      0.272      -0.489       1.695
==============================================================================
Omnibus:                       15.177   Durbin-Watson:                   1.990
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               28.147
Skew:                          -0.825   Prob(JB):                     7.73e-07
Kurtosis:                       6.284   Cond. No.                         5.72
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            GLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.276
Model:                            GLS   Adj. R-squared:                  0.261
Method:                 Least Squares   F-statistic:                     18.33
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           8.83e-05
Time:                        20:20:42   Log-Likelihood:                -191.29
No. Observations:                  50   AIC:                             386.6
Df Residuals:                      48   BIC:                             390.4
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.1989      2.281     -0.087      0.931      -4.785       4.387
x1            -1.3258      0.310     -4.280      0.000      -1.949      -0.703
==============================================================================
Omnibus:                       15.890   Durbin-Watson:                   2.001
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               31.492
Skew:                           0.837   Prob(JB):                     1.45e-07
Kurtosis:                       6.509   Cond. No.                         10.9
==============================================================================

Warnings:
```

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
In [14]: #Checking ols,gls performance on data without outliers
         def mse(actual, predicted):
             return ((actual - predicted)**2).mean()

         def RSQ(actual, predicted):
             y_bar=actual.mean()
             re=actual-predicted
             NUMM=sum(np.square(re))
             DOMM=sum(np.square(actual-y_bar))
             RESQ=1-(NUMM/DOMM)
             return(RESQ)

         dataP_nout={}
         ols_result_nout={}
         data_nout={}
         data_nout=data
         resp_nout={}
         resp_nout=resp

         data_nout[0] = data[0][data[0] != 4.0820779]
         resp_nout[0] = resp[0][resp[0] != -27.00803487]

         data_nout[1] = data[1][data[1] != -8.640359541]
         resp_nout[1] = resp[1][resp[1] != 77.64225929]

         data_nout[2] = data[2][data[2] != 13.60924464]
         resp_nout[2] = resp[2][resp[2] != 78.28113593]



         for i in range(0,5):
             dataP_nout[i]=sm.add_constant(data_nout[i])
             ols_result_nout[i] = sm.OLS(resp_nout[i],dataP_nout[i] ).fit()
             print(ols_result_nout[i].summary())


         Y_ols_nout={}
         for i in range (0,5):
             Y_ols_nout[i]=data[i]
             Y_ols_nout[i]=Y_ols_nout[i]*ols_result_nout[i].params[1]+ols_result_nout[i].params[
             print(ols_result_nout[i])
             print(RSQ(resp[i],Y_ols_nout[i]))


         ols_resid_nout={}
```

```python
resid_fit_nout={}
rho_nout={}
for i in range(0,5):
    ols_resid_nout[i] = ols_result_nout[i].resid
    resid_fit_nout[i] = sm.OLS(ols_resid_nout[i][1:], sm.add_constant(ols_resid_nout[i]
    rho_nout[i] = resid_fit_nout[i].params[1]

from scipy.linalg import toeplitz
order_nout={}
sigma_nout={}
gls_model_nout={}
gls_results_nout={}

toeplitz(range(5))
for i in range(0,5):
    order_nout[i] = toeplitz(range(len(ols_resid_nout[i])))
    sigma_nout[i] = rho_nout[i]**order_nout[i]
    gls_model_nout[i] = sm.GLS(resp_nout[i],dataP_nout[i], sigma=sigma_nout[i])
    gls_results_nout[i] = gls_model_nout[i].fit()
    print(gls_results_nout[i].summary())
    #print(gls_results[i].params)

Y_gls_nout={}
for i in range (0,5):
    Y_gls_nout[i]=data[i]
    Y_gls_nout[i]=Y_gls_nout[i]*gls_results_nout[i].params[1]+gls_results_nout[i].param
    print(gls_results_nout[i])
    print(RSQ(resp[i],Y_gls_nout[i]))
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.293
Model:                            OLS   Adj. R-squared:                  0.286
Method:                 Least Squares   F-statistic:                     40.26
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           7.10e-09
Time:                        20:21:58   Log-Likelihood:                -285.50
No. Observations:                  99   AIC:                             575.0
Df Residuals:                      97   BIC:                             580.2
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0129      0.440      0.029      0.977      -0.860       0.886
x1             1.0533      0.166      6.345      0.000       0.724       1.383
==============================================================================
Omnibus:                       18.485   Durbin-Watson:                   2.032
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               33.236
```

```
Skew:                           0.751   Prob(JB):                     6.07e-08
Kurtosis:                       5.409   Cond. No.                         2.66
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.531
Model:                            OLS   Adj. R-squared:                  0.526
Method:                 Least Squares   F-statistic:                     109.9
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           1.22e-17
Time:                        20:21:58   Log-Likelihood:                -391.23
No. Observations:                  99   AIC:                             786.5
Df Residuals:                      97   BIC:                             791.6
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -1.3455      1.280     -1.051      0.296      -3.887       1.196
x1             1.3726      0.131     10.483      0.000       1.113       1.632
==============================================================================
Omnibus:                        9.310   Durbin-Watson:                   2.257
Prob(Omnibus):                  0.010   Jarque-Bera (JB):               21.734
Skew:                          -0.057   Prob(JB):                     1.91e-05
Kurtosis:                       5.293   Cond. No.                         9.80
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.277
Model:                            OLS   Adj. R-squared:                  0.261
Method:                 Least Squares   F-statistic:                     17.97
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           0.000104
Time:                        20:21:58   Log-Likelihood:                -195.41
No. Observations:                  49   AIC:                             394.8
Df Residuals:                      47   BIC:                             398.6
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.3736      1.908      1.244      0.220      -1.465       6.212
x1            -0.7353      0.173     -4.239      0.000      -1.084      -0.386
==============================================================================
```

```
Omnibus:                       16.725   Durbin-Watson:                   2.066
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               24.690
Skew:                           1.081   Prob(JB):                     4.35e-06
Kurtosis:                       5.724   Cond. No.                         11.0
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.028
Model:                            OLS   Adj. R-squared:                  0.007
Method:                 Least Squares   F-statistic:                     1.362
Date:                Sun, 10 Mar 2019   Prob (F-statistic):              0.249
Time:                        20:21:58   Log-Likelihood:                -152.15
No. Observations:                  50   AIC:                             308.3
Df Residuals:                      48   BIC:                             312.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.1481      1.443      0.796      0.430      -1.753       4.049
x1             0.6334      0.543      1.167      0.249      -0.458       1.725
==============================================================================
Omnibus:                       15.626   Durbin-Watson:                   1.936
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               29.405
Skew:                          -0.848   Prob(JB):                     4.12e-07
Kurtosis:                       6.352   Cond. No.                         5.81
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.281
Model:                            OLS   Adj. R-squared:                  0.266
Method:                 Least Squares   F-statistic:                     18.72
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           7.63e-05
Time:                        20:21:58   Log-Likelihood:                -191.30
No. Observations:                  50   AIC:                             386.6
Df Residuals:                      48   BIC:                             390.4
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.2413      2.302     -0.105      0.917      -4.870       4.387
```

```
x1            -1.3331      0.308     -4.327      0.000      -1.953      -0.714
==============================================================================
Omnibus:                       15.549   Durbin-Watson:                   2.048
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               29.921
Skew:                           0.830   Prob(JB):                     3.18e-07
Kurtosis:                       6.407   Cond. No.                         10.8
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x0000024687416588>
0.2933129059724171
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x00000246874165F8>
0.531163930110313
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x0000024687233BE0>
0.27658535243422755
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x00000246873F9320>
0.02759473593477113
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x000002468740CDD8>
0.28057579271914335
                            GLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.294
Model:                            GLS   Adj. R-squared:                  0.287
Method:                 Least Squares   F-statistic:                     40.37
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           6.83e-09
Time:                        20:21:58   Log-Likelihood:                -285.48
No. Observations:                  99   AIC:                             575.0
Df Residuals:                      97   BIC:                             580.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0119      0.432      0.028      0.978      -0.845       0.869
x1             1.0538      0.166      6.354      0.000       0.725       1.383
==============================================================================
Omnibus:                       18.098   Durbin-Watson:                   1.996
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               31.992
Skew:                           0.742   Prob(JB):                     1.13e-07
Kurtosis:                       5.357   Cond. No.                         2.61
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            GLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.561
```

```
Model:                            GLS   Adj. R-squared:                  0.556
Method:                 Least Squares   F-statistic:                     123.8
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           5.02e-19
Time:                        20:21:58   Log-Likelihood:                -390.21
No. Observations:                  99   AIC:                             784.4
Df Residuals:                      97   BIC:                             789.6
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -1.3971      1.115     -1.253      0.213      -3.610       0.816
x1             1.4158      0.127     11.128      0.000       1.163       1.668
==============================================================================
Omnibus:                        8.769   Durbin-Watson:                   2.002
Prob(Omnibus):                  0.012   Jarque-Bera (JB):               19.543
Skew:                          -0.009   Prob(JB):                     5.71e-05
Kurtosis:                       5.177   Cond. No.                         8.78
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            GLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.282
Model:                            GLS   Adj. R-squared:                  0.267
Method:                 Least Squares   F-statistic:                     18.45
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           8.68e-05
Time:                        20:21:58   Log-Likelihood:                -195.37
No. Observations:                  49   AIC:                             394.7
Df Residuals:                      47   BIC:                             398.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.3556      1.832      1.286      0.205      -1.330       6.041
x1            -0.7422      0.173     -4.295      0.000      -1.090      -0.395
==============================================================================
Omnibus:                       16.298   Durbin-Watson:                   1.976
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               24.017
Skew:                           1.051   Prob(JB):                     6.09e-06
Kurtosis:                       5.711   Cond. No.                         10.6
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            GLS Regression Results
```

```
==============================================================================
Dep. Variable:                      y   R-squared:                       0.025
Model:                            GLS   Adj. R-squared:                  0.005
Method:                 Least Squares   F-statistic:                     1.243
Date:                Sun, 10 Mar 2019   Prob (F-statistic):              0.270
Time:                        20:21:58   Log-Likelihood:                -152.12
No. Observations:                  50   AIC:                             308.2
Df Residuals:                      48   BIC:                             312.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.2190      1.454      0.839      0.406      -1.704       4.142
x1             0.6028      0.543      1.110      0.272      -0.489       1.695
==============================================================================
Omnibus:                       15.177   Durbin-Watson:                   1.990
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               28.147
Skew:                          -0.825   Prob(JB):                     7.73e-07
Kurtosis:                       6.284   Cond. No.                         5.72
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                            GLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.276
Model:                            GLS   Adj. R-squared:                  0.261
Method:                 Least Squares   F-statistic:                     18.33
Date:                Sun, 10 Mar 2019   Prob (F-statistic):           8.83e-05
Time:                        20:21:58   Log-Likelihood:                -191.29
No. Observations:                  50   AIC:                             386.6
Df Residuals:                      48   BIC:                             390.4
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.1989      2.281     -0.087      0.931      -4.785       4.387
x1            -1.3258      0.310     -4.280      0.000      -1.949      -0.703
==============================================================================
Omnibus:                       15.890   Durbin-Watson:                   2.001
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               31.492
Skew:                           0.837   Prob(JB):                     1.45e-07
Kurtosis:                       6.509   Cond. No.                         10.9
==============================================================================

Warnings:
```

```
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x0000024687233128>
0.2933128083201585
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x000002468740CE80>
0.5306346544176397
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x000002468724AFD0>
0.27656085293279886
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x00000246873F90F0>
0.02753049977689004
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x00000246873F9E10>
0.28056735892389884
```

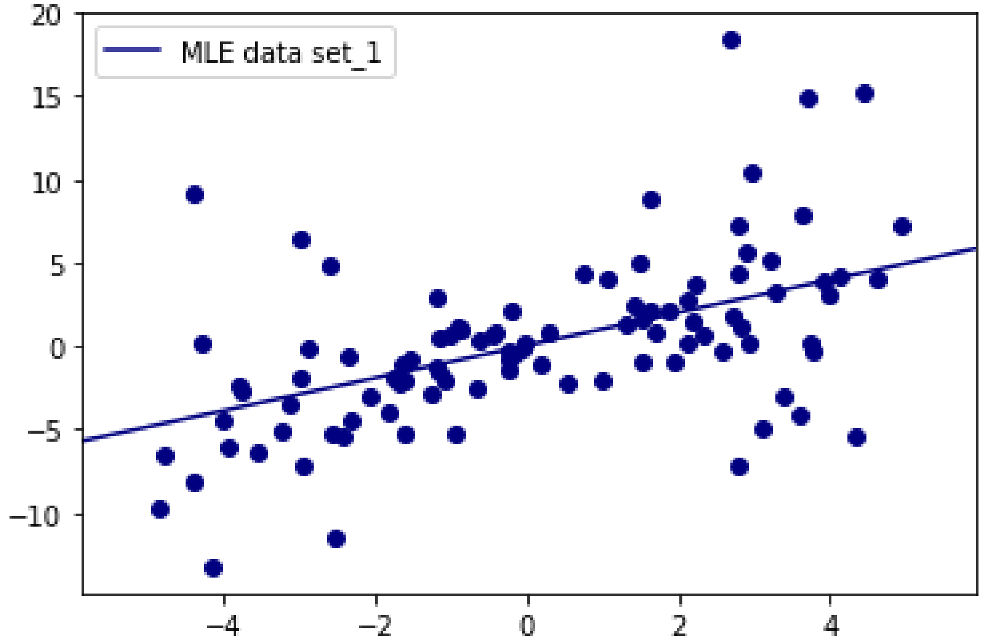
```
In [15]: #MLE with a noise dependent to x
         from scipy.optimize import minimize
         import math
         def myfunc(params):
             # print(params)  # <-- you'll see that params is a NumPy array
             a, b, s0, s1 = params # <-- for readability you may wish to assign names to the com
             X=np.loadtxt('X_5.txt')
             Y=np.loadtxt('Y_5.txt')
             ss=np.ones((50))
             ress=np.zeros((50))
             kapak=np.zeros((50))
             #print (ress)
             for i in range(0,50):
                 ss[i]=((X[i]*s1)**2)+(s0**2)
                 ress[i]=((Y[i]-b*X[i]-a)**2)/(2*ss[i])
                 #print(ss)
                 kapak[i]=math.log(ss[i])

             NLL= 0.5*np.sum(kapak)+ np.sum(ress)
             return NLL


         myresult= minimize(myfunc, [1 ,1 ,1 ,1],method='BFGS')
         print(myresult)
```

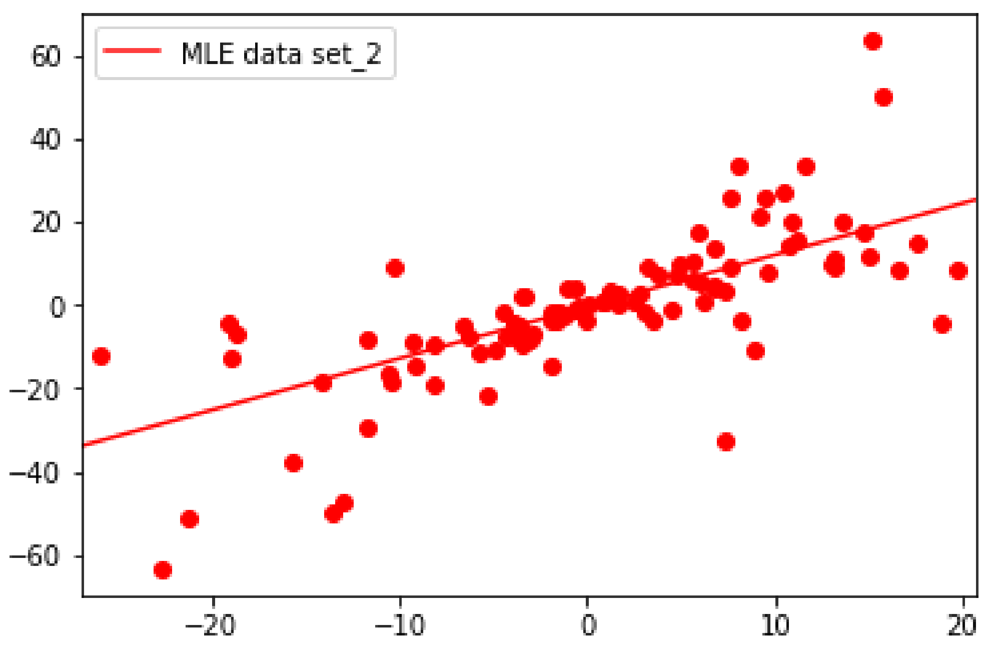
```
      fun: 114.81521007647805
 hess_inv: array([[ 0.12640253,  0.01813645,  0.01810048, -0.00426204],
       [ 0.01813645,  0.04218741, -0.00018548, -0.00023481],
       [ 0.01810048, -0.00018548,  0.10419203, -0.00919975],
       [-0.00426204, -0.00023481, -0.00919975,  0.02230889]])
      jac: array([9.53674316e-07, 5.72204590e-06, 1.90734863e-06, 4.76837158e-06])
  message: 'Optimization terminated successfully.'
     nfev: 114
      nit: 12
     njev: 19
```

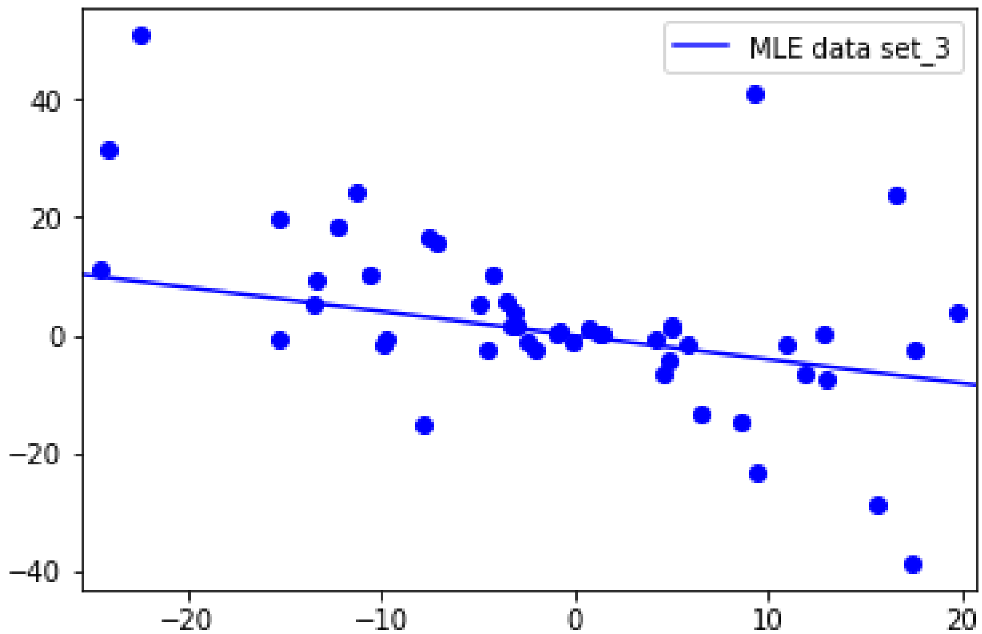
```
   status: 0
  success: True
        x: array([ 0.23780701, -0.85762322,  0.81847457,  1.34786082])
```

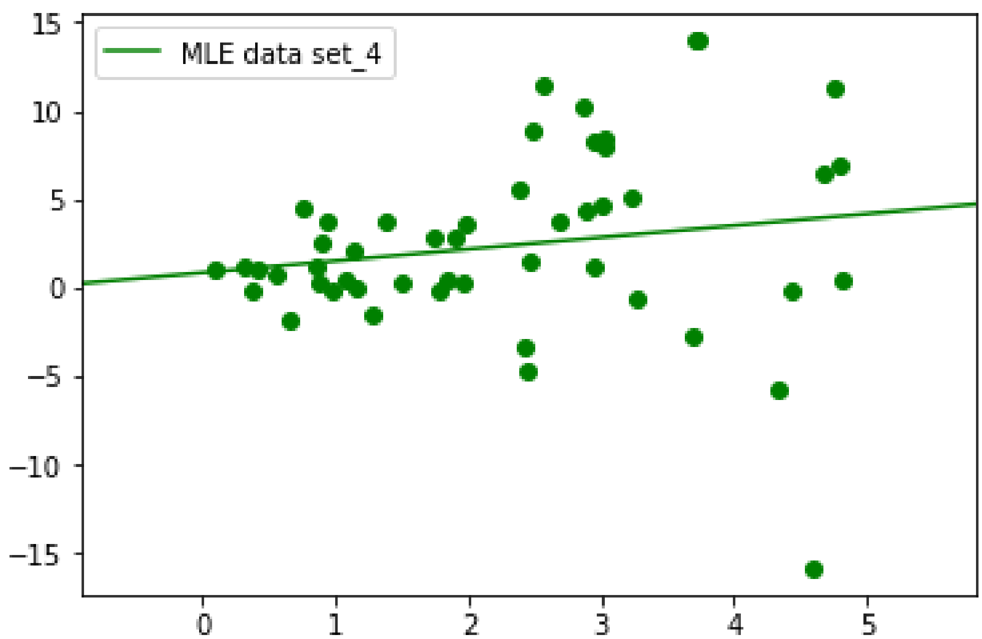
```
In [19]: MLE_coef=([0.05307253, 0.97987193], [-0.43664784,  1.24209258], [-0.03350694, -0.405626
         #print(MLE_coef[0][0])
         Y_MLX={}
         for i in range (0,5):
             Y_MLX[i]=data[i]
             Y_MLX[i]=Y_MLX[i]*MLE_coef[i][1]+MLE_coef[i][0]
             print('MLE coef for data_set_'+str(i+1),MLE_coef[i])
             print('MLE Rsq data_set_'+str(i+1),(RSQ(resp[i],Y_MLX[i])))

         datas={}
         Datas={}
         for i in range(0,5):
             datas[i]= np.linspace(MINN[i], MAXX[i], 20)
             Datas[i]= sm.add_constant(datas[i])
             fig, ax = plt.subplots()
             ax.scatter(data[i], resp[i], color=C[i]);
             ax.plot(datas[i], datas[i]*MLE_coef[i][1]+MLE_coef[i][0], color=C[i], label='MLE mo
             ax.set_xlim(MINN[i], MAXX[i]);
             ax.legend(loc='upper right');
             ax.legend(['MLE data set_'+str(i+1)])
```

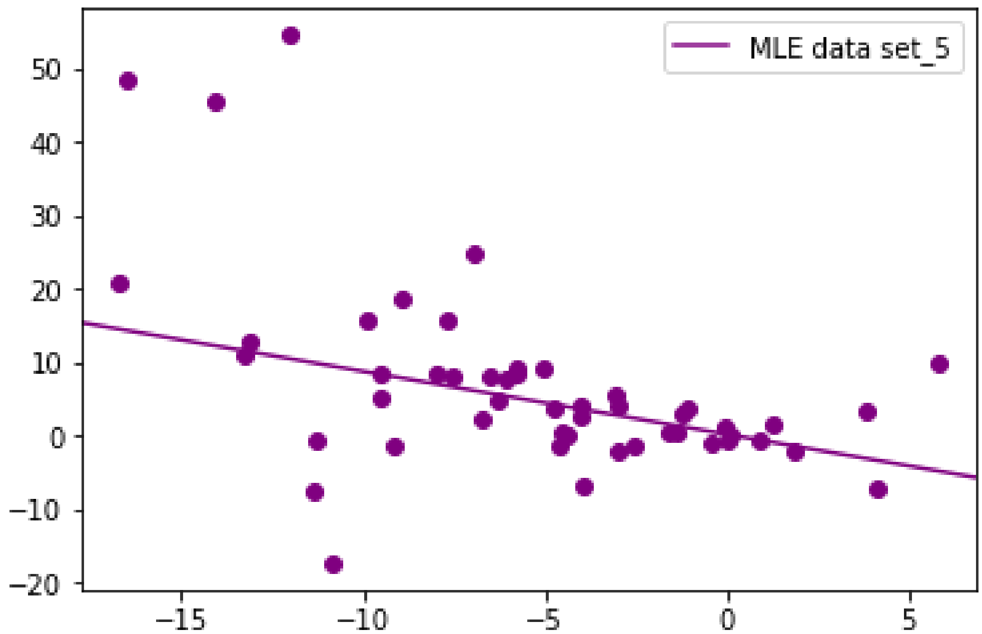
```
MLE coef for data_set_1 [0.05307253, 0.97987193]
MLE Rsq data_set_1 0.29185398557586284
MLE coef for data_set_2 [-0.43664784, 1.24209258]
MLE Rsq data_set_2 0.5243006087349533
MLE coef for data_set_3 [-0.03350694, -0.40562675]
MLE Rsq data_set_3 0.19137528869671872
MLE coef for data_set_4 [0.848970404, 0.665350423]
MLE Rsq data_set_4 0.025596373566421526
MLE coef for data_set_5 [0.23780701, -0.85762322]
MLE Rsq data_set_5 0.21984216965618752
```


MLE data set_1
20
15
10
5
0
−5
−10
−4
−2
0
2
4


MLE data set_2
60
40
20
0
−20
−40
−60
−20
−10
0
10
20


MLE data set_3
40
20
0
−20
−40
−20
−10
0
10
20

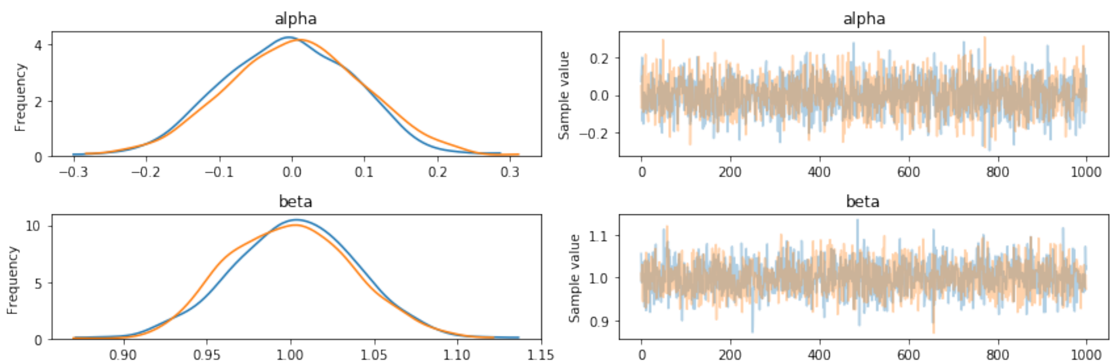

MLE data set_4
15
10
5
0
−5
−10
−15
0
1
2
3
4
5

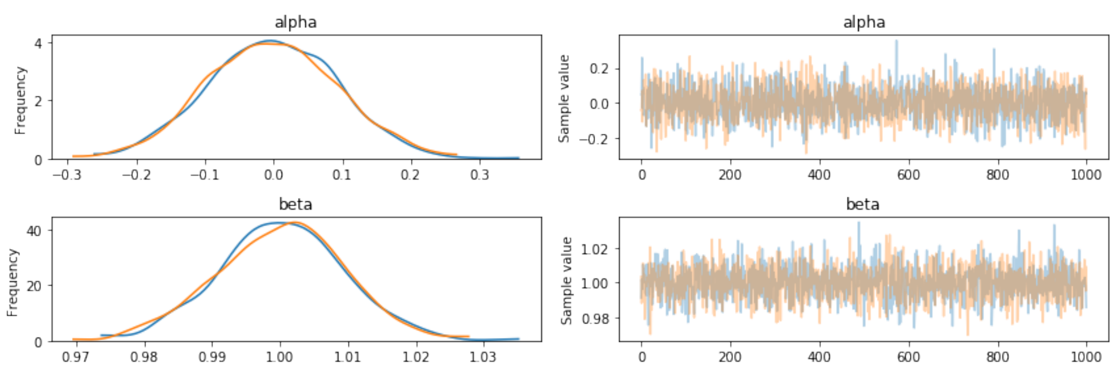


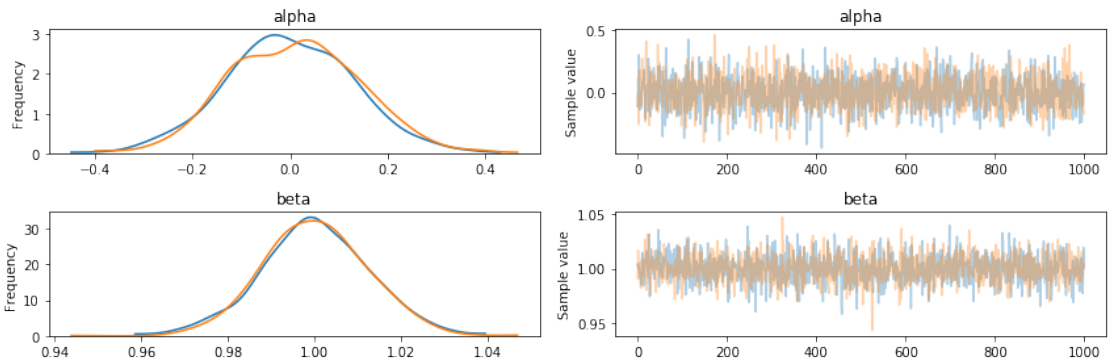

```
In [17]: #Baysian Modelling
         basic_model={}
         y_hat={}
         y_obser={}
         step={}
         trace={}
         trace_={}
         for i in range(0,5):
             basic_model[i]= pm.Model()
             # Regression coefficients
             with basic_model[i]:
                 alpha = pm.Uniform('alpha', -100, 100)
                 beta = pm.Uniform('beta', -100, 100)

             # Expected value
                 y_hat[i] =  data[i]* beta+ alpha
             # Observations with t-distributed error
                 y_obser[i] = pm.StudentT('y_obs[i]', nu=5, mu=y_hat[i], observed=data[i])
                 step[i] = pm.NUTS()
                 trace_[i] = pm.sample(3000, step[i])
                 burn = 1000
                 thin = 2
                 trace[i] = trace_[i][burn::thin]
                 pm.plots.traceplot(trace[i]);
```

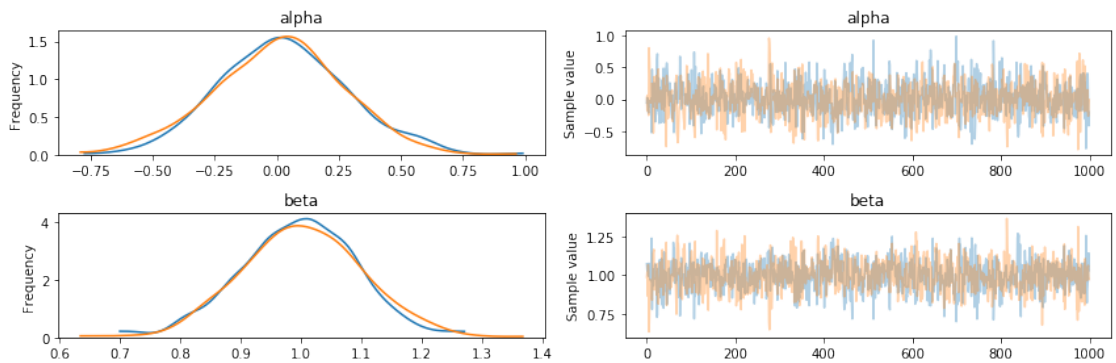
```
Multiprocess sampling (2 chains in 2 jobs)
```

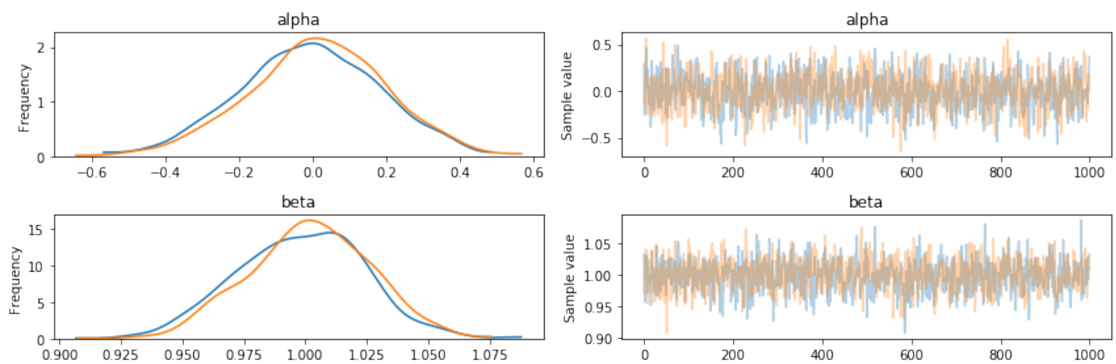
```
NUTS: [beta, alpha]
Sampling 2 chains: 100%|[U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588
Multiprocess sampling (2 chains in 2 jobs)
NUTS: [beta, alpha]
Sampling 2 chains: 100%|[U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588
Multiprocess sampling (2 chains in 2 jobs)
NUTS: [beta, alpha]
Sampling 2 chains: 100%|[U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588
Multiprocess sampling (2 chains in 2 jobs)
NUTS: [beta, alpha]
Sampling 2 chains: 100%|[U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588
The acceptance probability does not match the target. It is 0.8906291527658613, but should be cl
Multiprocess sampling (2 chains in 2 jobs)
NUTS: [beta, alpha]
Sampling 2 chains: 100%|[U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588
The acceptance probability does not match the target. It is 0.882617150435636, but should be clo
The acceptance probability does not match the target. It is 0.8996294521462941, but should be cl
```







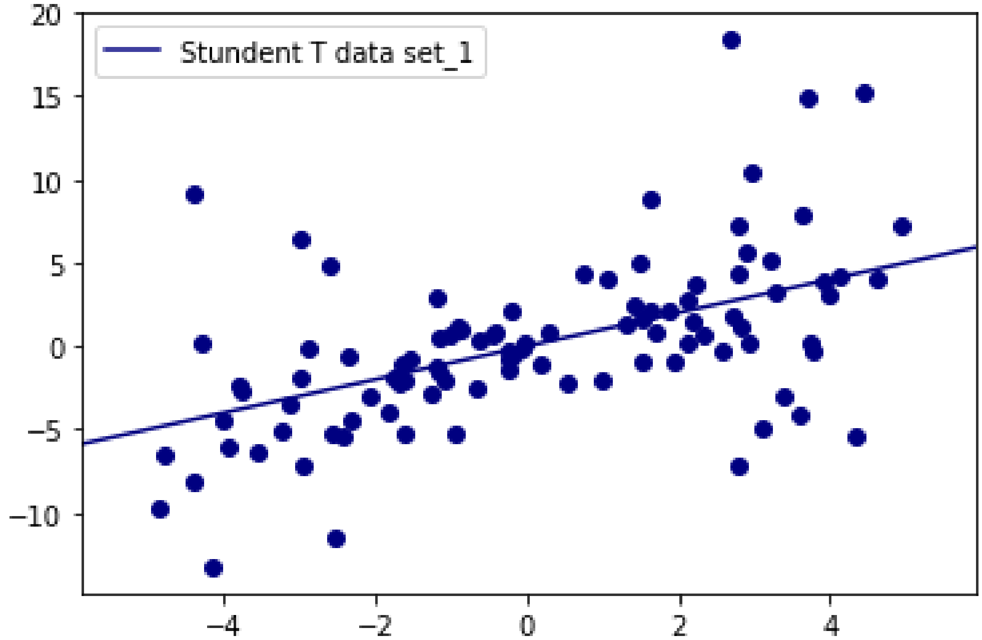


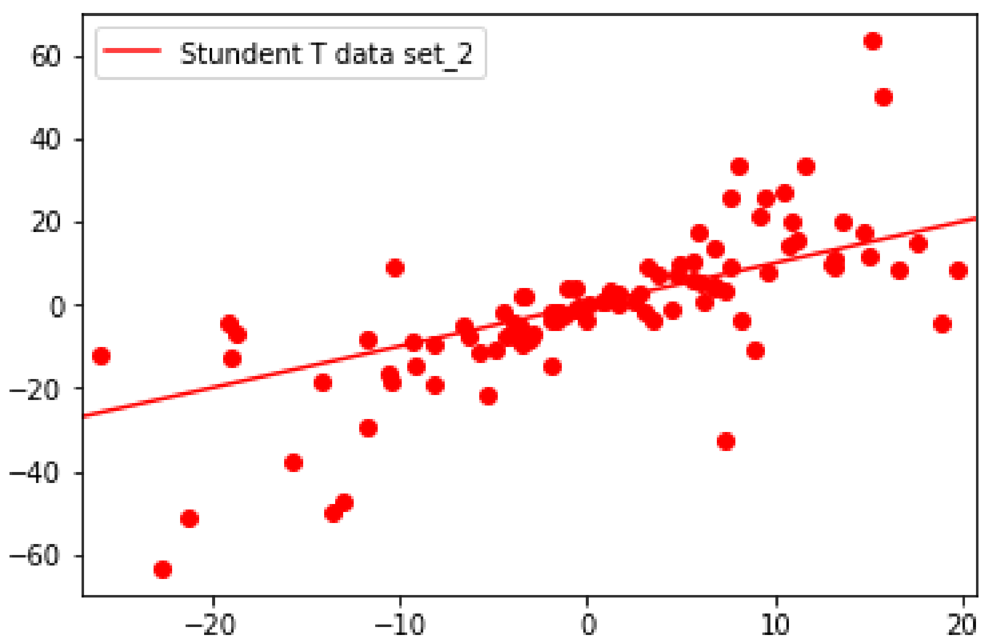


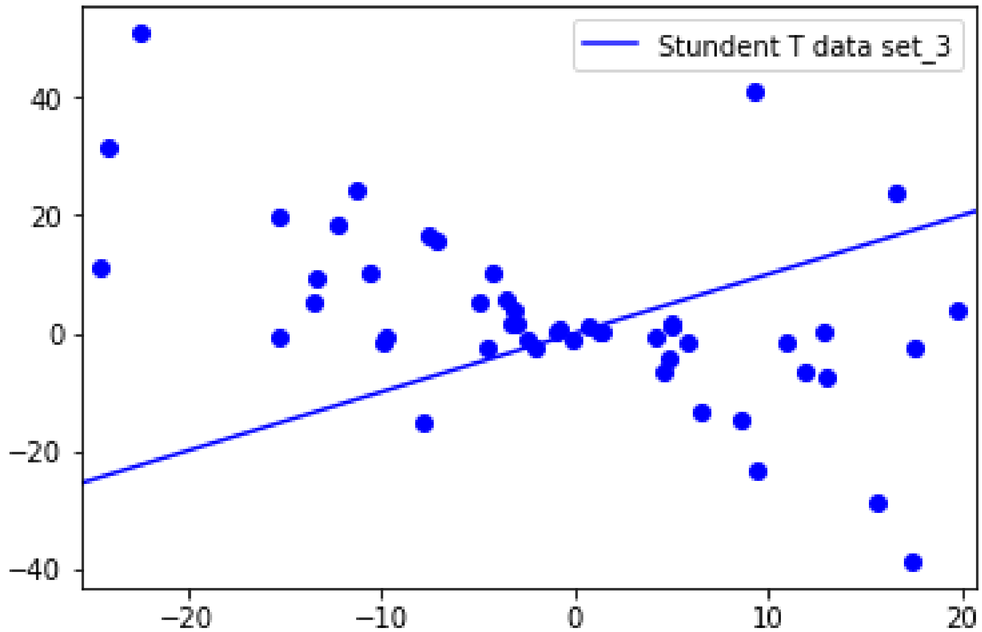

```
In [23]: alpha={}
         beta={}
         for i in range (0,5):
             alpha[i] = trace[i]['alpha'].mean()
             beta[i] = trace[i]['beta'].mean()
```

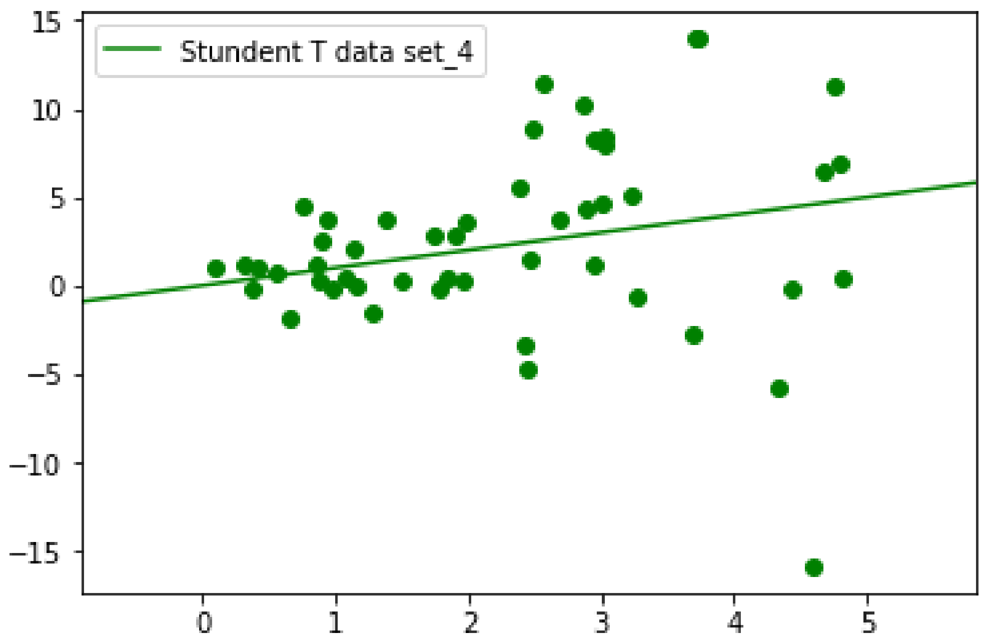
```
Y_Rob={}
for i in range (0,5):
    Y_Rob[i]=data[i]
    Y_Rob[i]=Y_Rob[i]*beta[i]+alpha[i]
    print('Student-T coef for data_set_'+str(i+1),beta[i],alpha[i])
    print('Rsq for data_set_'+str(i+1),(RSQ(resp[i],Y_Rob[i])))

datas={}
Datas={}
for i in range(0,5):
    datas[i]= np.linspace(MINN[i], MAXX[i], 20)
    Datas[i]= sm.add_constant(datas[i])
    fig, ax = plt.subplots()
    ax.scatter(data[i], resp[i],color=C[i]);
    ax.plot(datas[i], datas[i]*beta[i]+alpha[i], color=C[i], label='Robust T-dist model
    ax.set_xlim(MINN[i], MAXX[i]);
    ax.legend(loc='upper right');
    ax.legend(['Stundent T data set_'+str(i+1)])
```

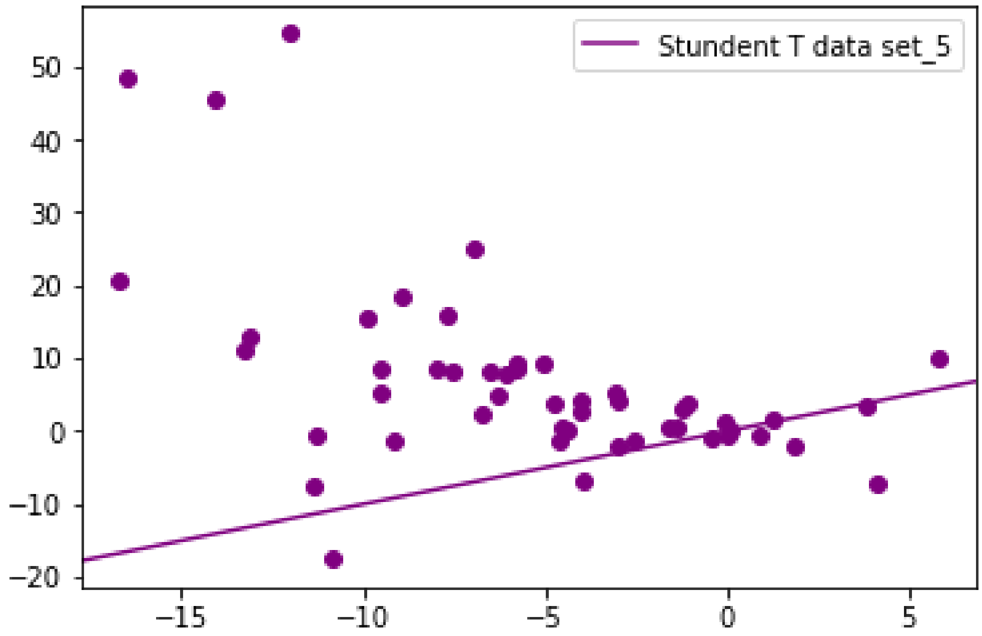
```
Student-T coef for data_set_1 1.0008706022442138 -0.0005895080337954823
Rsq for data_set_1 0.29257019114043736
Student-T coef for data_set_2 1.000082644975639 -0.002710147107545801
Rsq for data_set_2 0.4882526329546035
Student-T coef for data_set_3 1.000022650754439 0.0024073996467028065
Rsq for data_set_3 -1.3188258803371764
Student-T coef for data_set_4 0.9970683439054323 0.01498564398632196
Rsq for data_set_4 0.015098503877479241
Student-T coef for data_set_5 0.9994572175384412 -0.0018699430398291738
Rsq for data_set_5 -1.4576404973251789
```


Stundent T data set_1
20
15
10
5
0
−5
−10
−4
−2
0
2
4


Stundent T data set_2
60
40
20
0
−20
−40
−60
−20
−10
0
10
20


Stundent T data set_3
40
20
0
−20
−40
−20
−10
0
10
20


Stundent T data set_4
15
10
5
0
−5
−10
−15
0
1
2
3
4
5



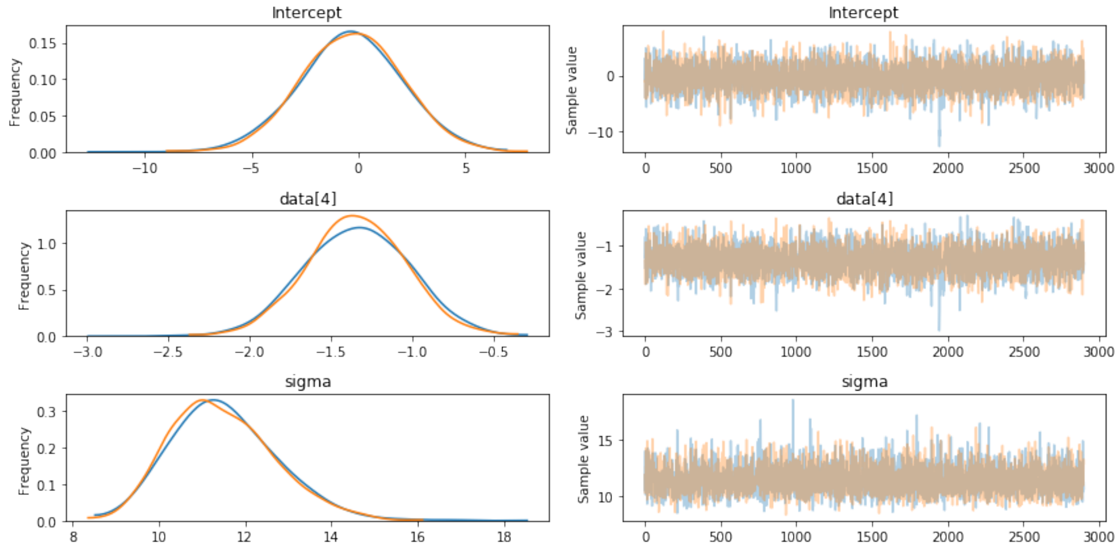

```
In [25]: #generating Half-Cauchy
         modell=pm.Model()
         with modell: # model specifications in PyMC3 are wrapped in a with-statement
             # Define priors
             sigma = pm.distributions.continuous.HalfCauchy('sigma', beta=10, testval=1.)
             intercept = pm.distributions.continuous.Normal('Intercept', 0, sd=20)
             x_coeff = pm.distributions.continuous.Normal('data[4]', 0, sd=20)

             # Define likelihood
             likelihood = pm.distributions.continuous.Normal('resp[4]', mu=intercept + x_coeff *
                                     sd=sigma, observed=resp[4])

             # Inference!
             trace = pm.sample(3000, cores=2) # draw 3000 posterior samples using NUTS sampling
```

```
Auto-assigning NUTS sampler...
Initializing NUTS using jitter+adapt_diag...
Multiprocess sampling (2 chains in 2 jobs)
NUTS: [data[4], Intercept, sigma]
Sampling 2 chains: 100%|[U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+2588][U+258
The acceptance probability does not match the target. It is 0.8869367135055707, but should be cl
```

```
In [26]: plt.figure(figsize=(7, 7))
         pm.traceplot(trace[100:])
         plt.tight_layout();
```

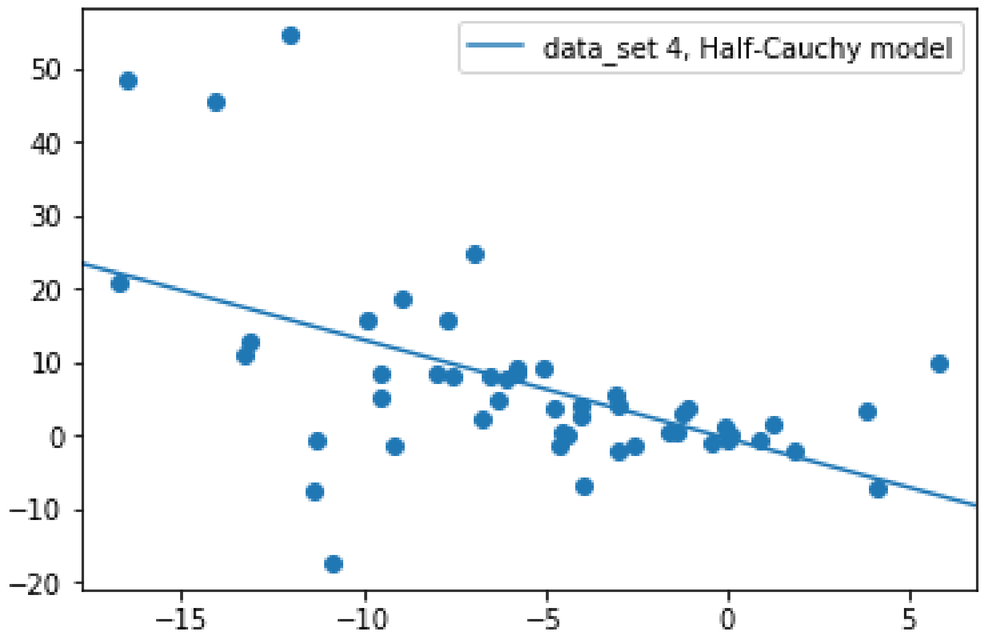
```
# with modell:
#     # specify glm and pass in data. The resulting linear model, its likelihood and
#     # and all its parameters are automatically added to our model.
#     glm.GLM.from_formula('y ~ x', data[4])
#     trace = pm.sample(3000, cores=2)
```

<Figure size 504x504 with 0 Axes>

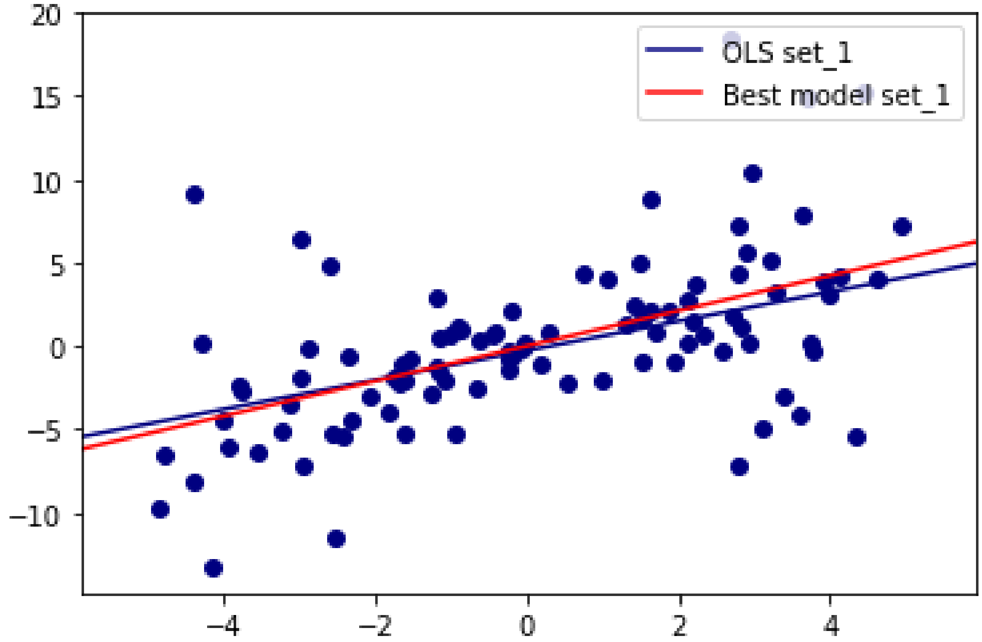


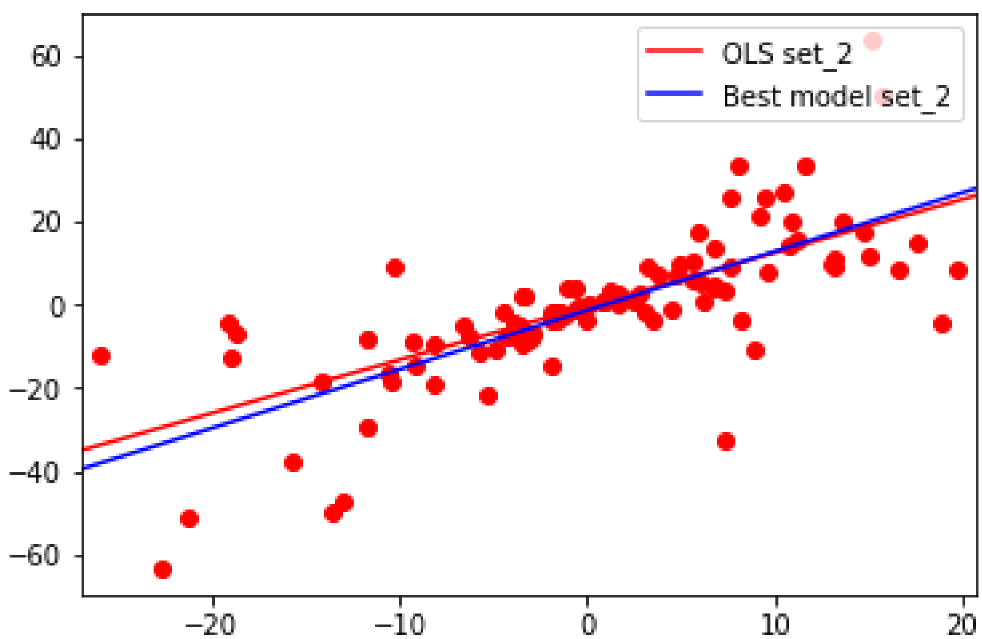

```
In [27]: alpha = trace['Intercept'].mean()
         beta = trace['data[4]'].mean()
         Y_Rob=data[4]
         Y_Rob=Y_Rob*beta+alpha
         print('coef for data_set_4',alpha,beta)
         print('Rsqr', RSQ(resp[4],Y_Rob))
         datas= np.linspace(MINN[4], MAXX[4], 20)
         Datas= sm.add_constant(datas)
         fig, ax = plt.subplots()
         ax.scatter(data[4], resp[4]);
         ax.plot(datas, datas*beta+alpha, label='data_set 4, Half-Cauchy model');
         ax.set_xlim(MINN[4], MAXX[4]);
         ax.legend(loc='upper right');
```

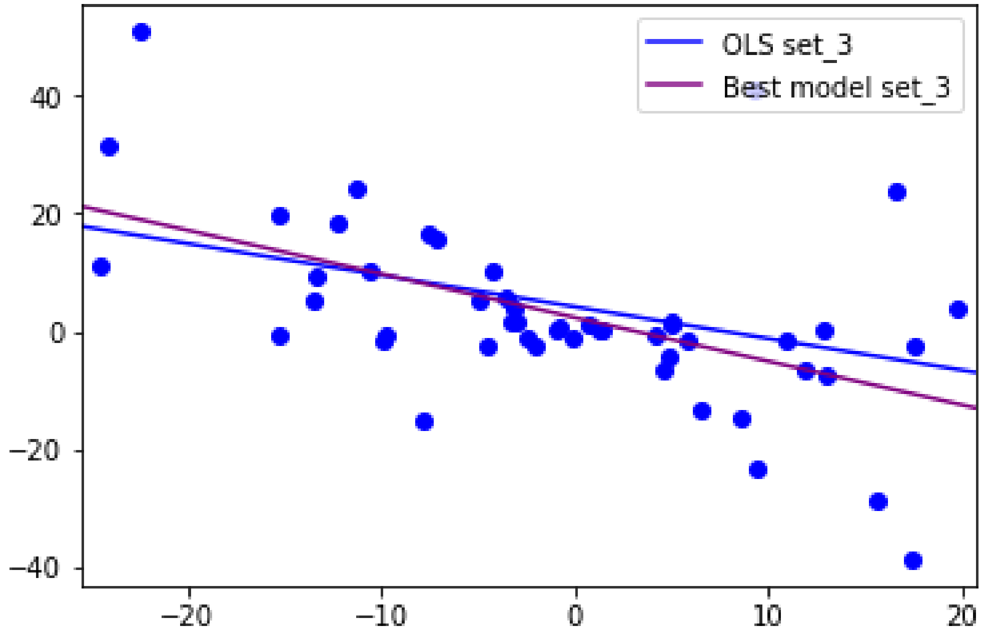
```
coef for data_set_4 -0.32369707443170476 -1.3409060717512427
Rsqr 0.28055659261033905
```

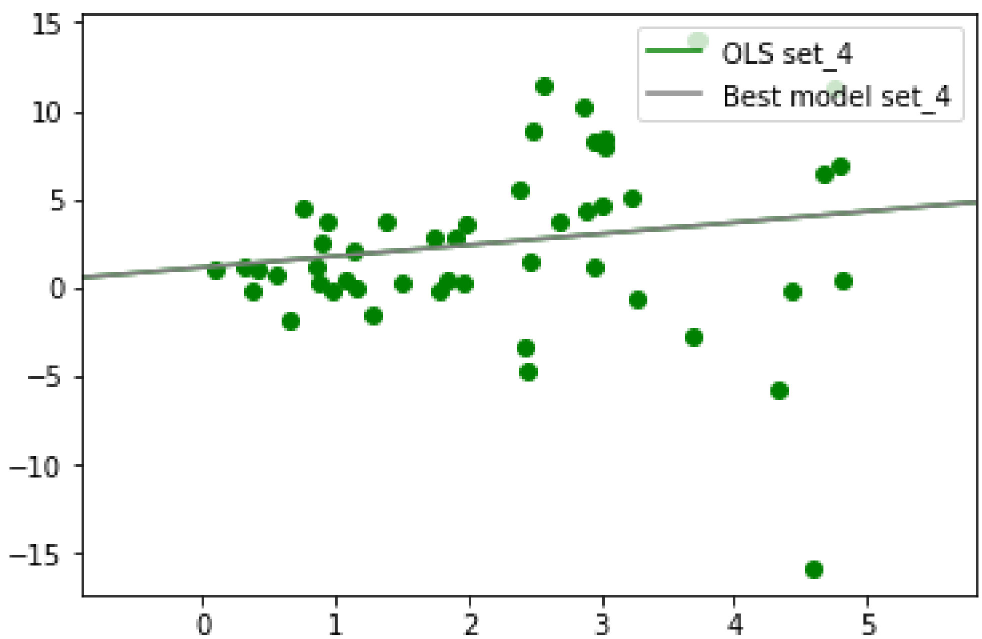


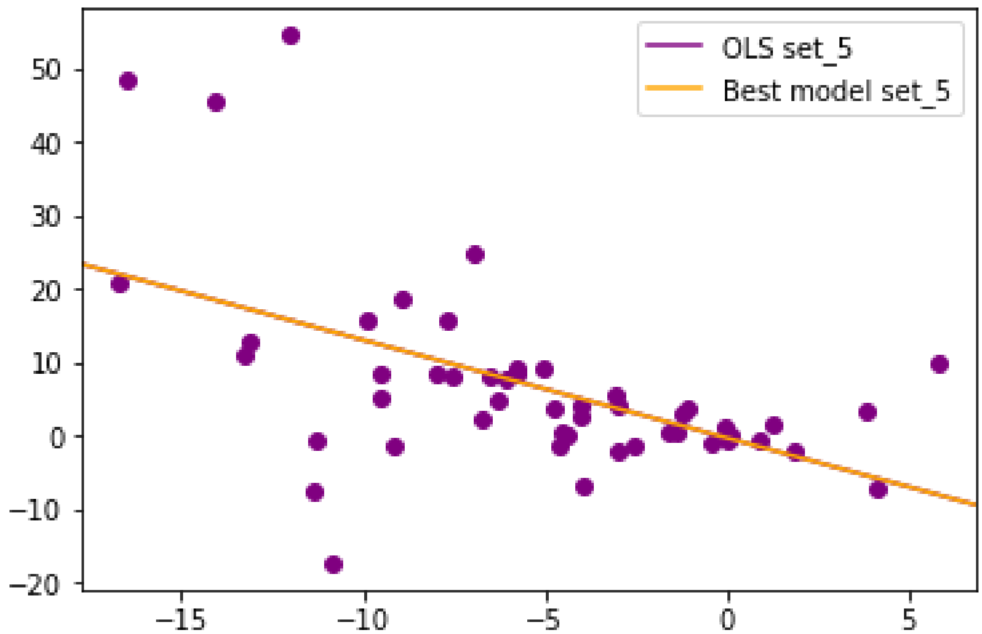

```
In [28]: datas={}
         Datas={}
         alphaa=[0.0119,-1.3971,2.3556,1.1481,-0.2413]
         DD=['red', 'blue', 'purple','gray','orange']
         betaa=[1.0538,1.4158,-0.7422,0.6334,-1.3331]
         for i in range(0,5):
             datas[i]= np.linspace(MINN[i], MAXX[i], 20)
             Datas[i]= sm.add_constant(datas[i])
             fig, ax = plt.subplots()
             ax.scatter(data[i], resp[i],color=C[i]);
             ax.plot(datas[i], ols_result[i].predict(Datas[i]), color=C[i],label='OLS set_'+str(
             ax.set_xlim(MINN[i], MAXX[i]);


             ax.plot(datas[i], alphaa[i] + betaa[i] * datas[i],color=DD[i], label='Best model se

             ax.legend(loc='upper right');
```


OLS set_1
Best model set_1
20
15
10
5
0
−5
−10
−4
−2
0
2
4


OLS set_2
Best model set_2
60
40
20
0
−20
−40
−60
−20
−10
0
10
20


OLS set_3
Best model set_3
40
20
0
−20
−40
−20
−10
0
10
20


OLS set_4
Best model set_4
15
10
5
0
−5
−10
−15
0
1
2
3
4
5




```
In [ ]:
```

```
In [ ]:
```