lab1_nsak

March 8, 2022

1 Lab 1 Norbert Sak (group 1a)

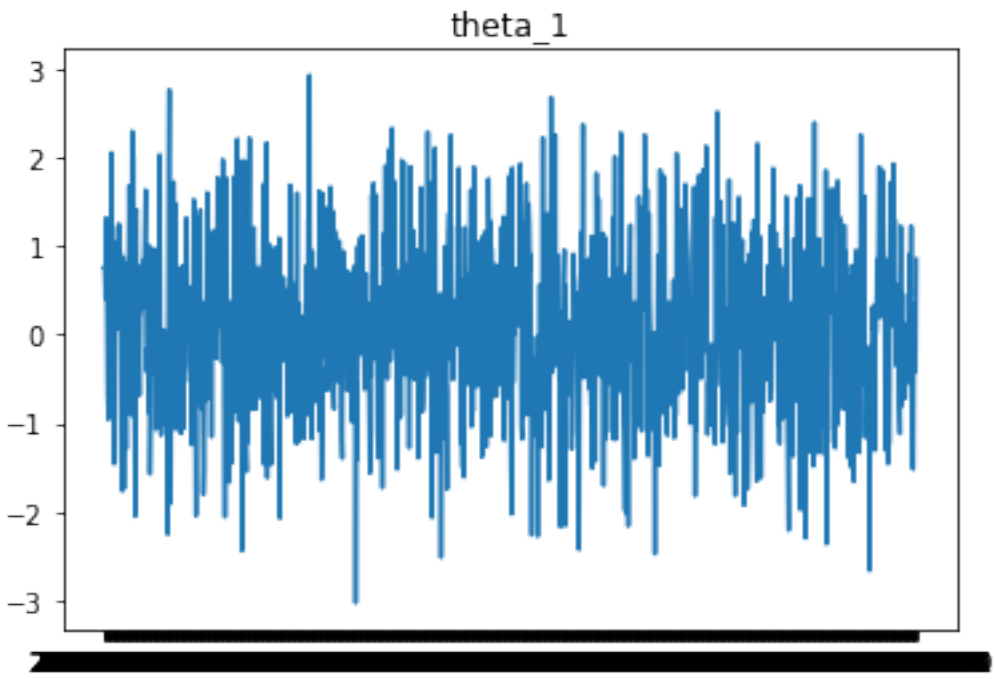
1.1 Excersise 1

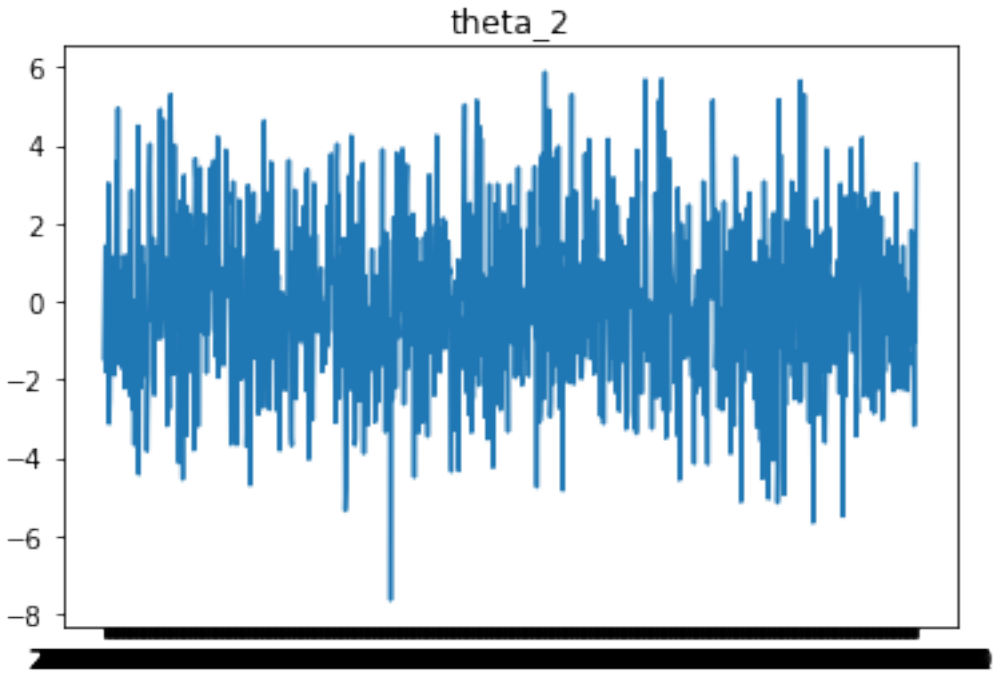
[ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("Data1.csv", index_col=0)

df.head()
```

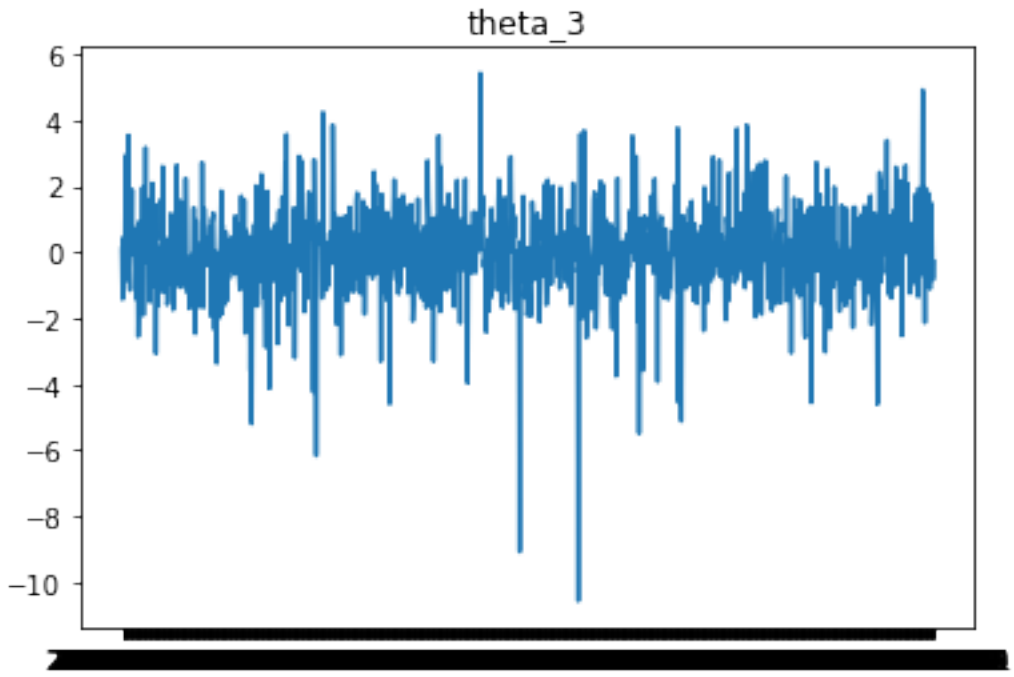
[ ]:

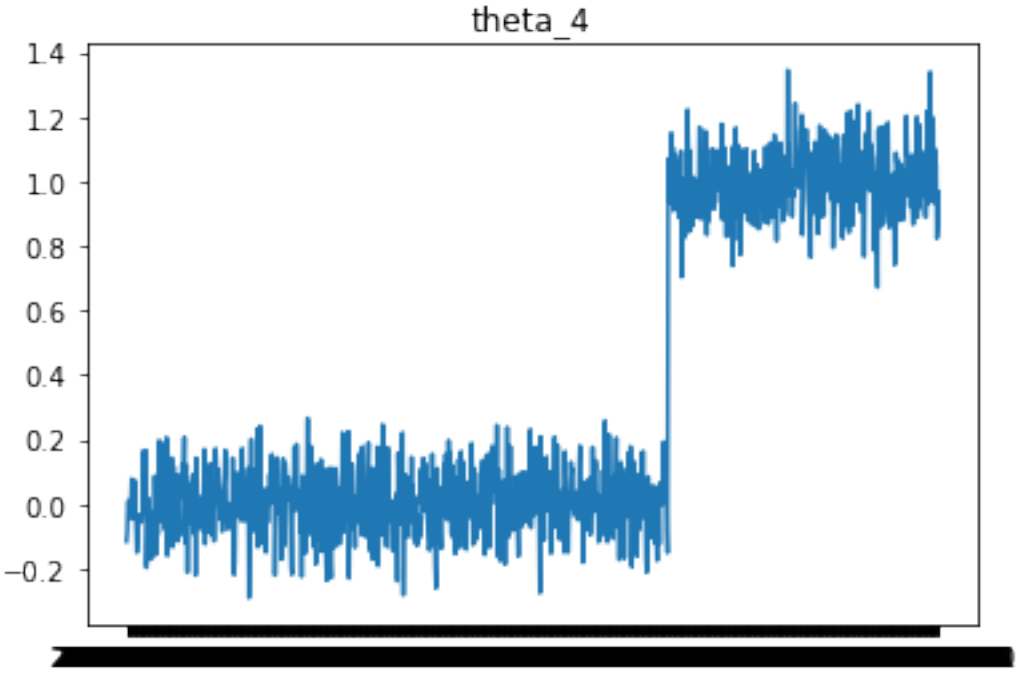
	theta_1	theta_2	theta_3	theta_4	theta_5	theta_6
2017-01-01	0.756936	-1.467790	0.096136	-0.115306	-0.447908	0.902579
2017-01-02	0.767089	0.185797	-1.428536	-0.086443	-0.954288	1.930909
2017-01-03	0.404544	1.415887	0.443466	0.000200	-0.892351	2.449691
2017-01-04	1.313957	-1.804471	-0.836986	0.011785	-1.012518	1.182085
2017-01-05	0.209862	1.315868	0.140993	-0.046473	-1.417092	1.742433

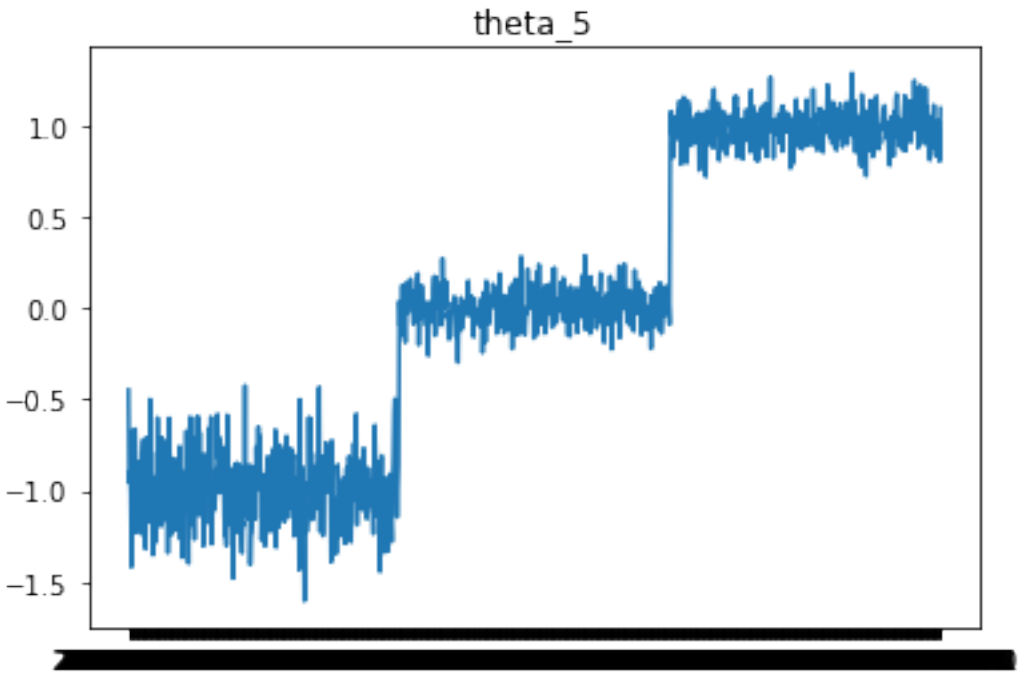
[ ]:
```
for i in df.columns:
    plt.plot(df.index, df[i])
    plt.title(i)
    plt.show()
```

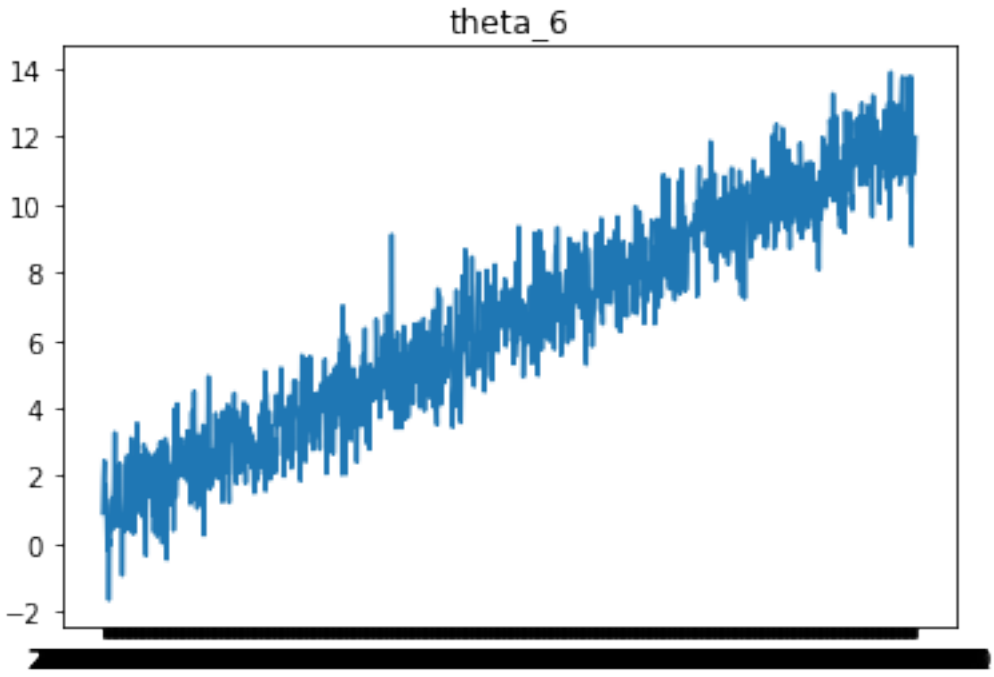

theta_1
3
2
1
0
−1
−2
−3


theta_2
6
4
2
0
−2
−4
−6
−8


theta_3
6
4
2
0
−2
−4
−6
−8
−10

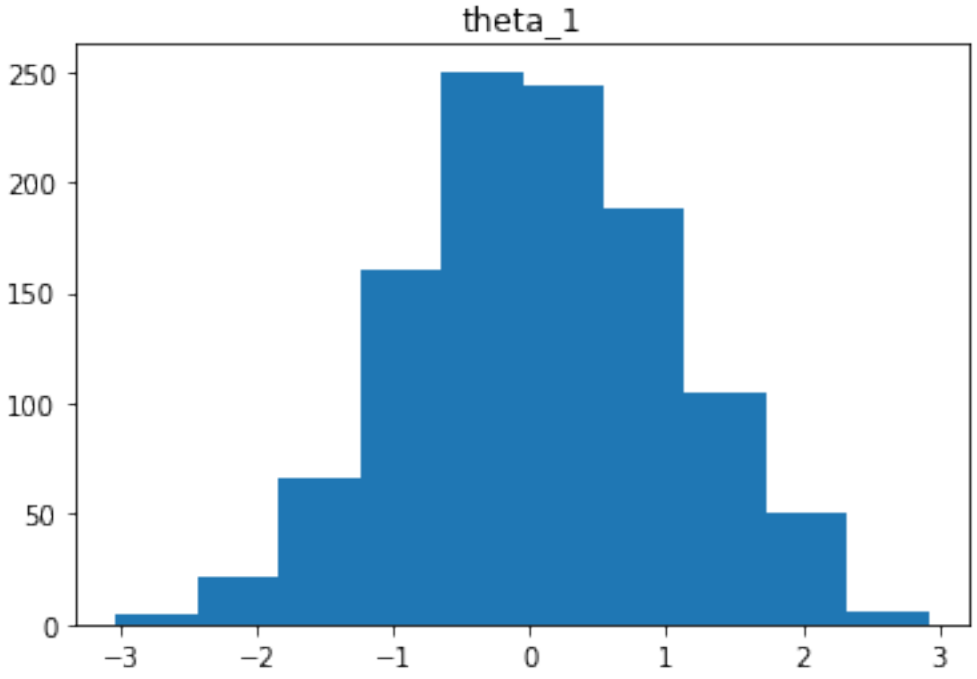

theta_4
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

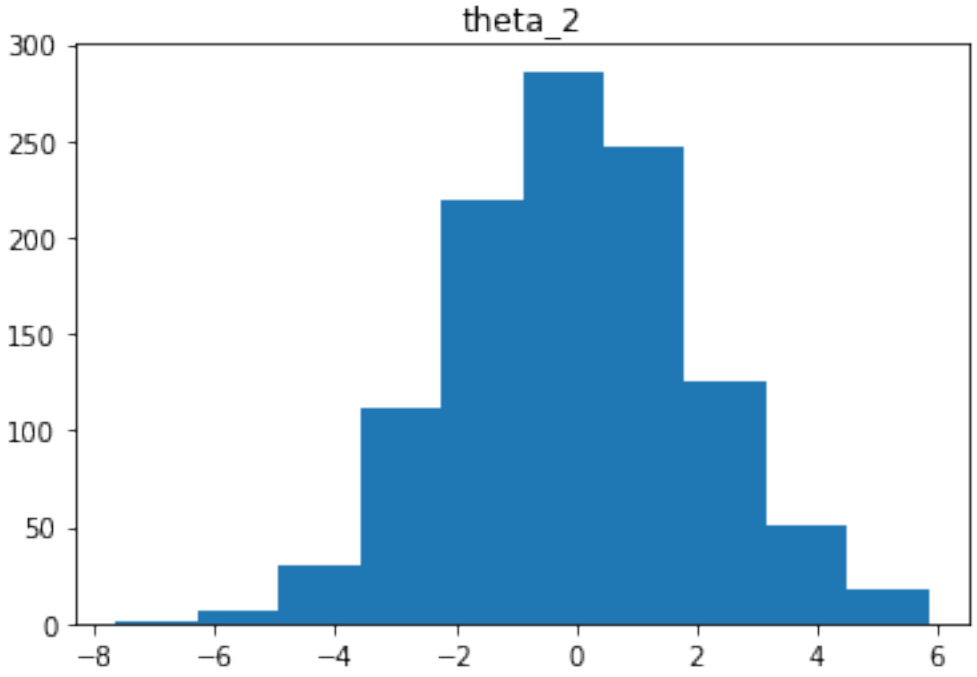


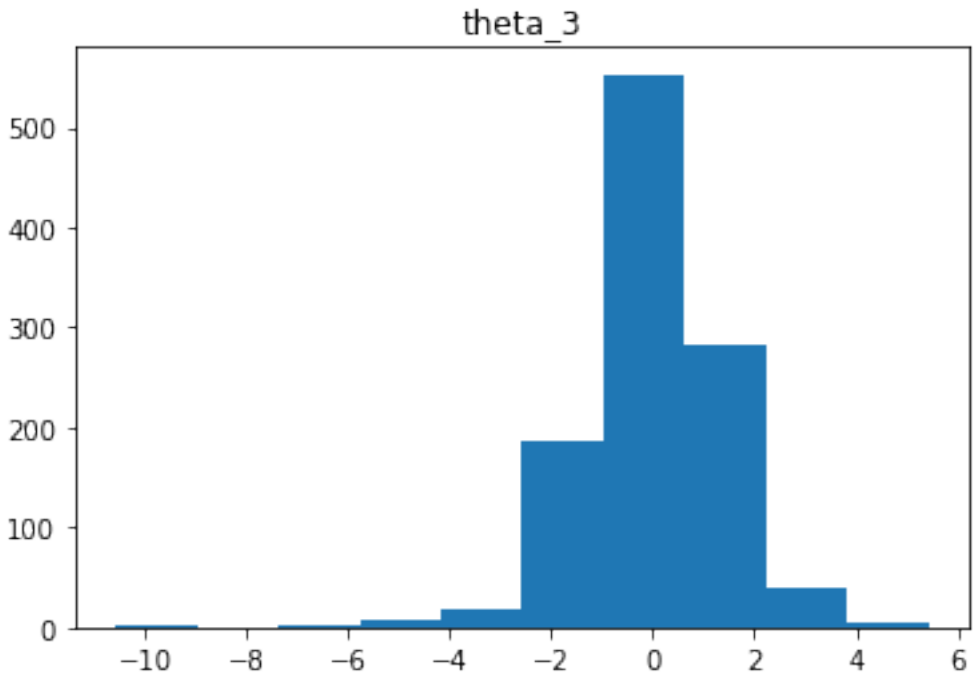


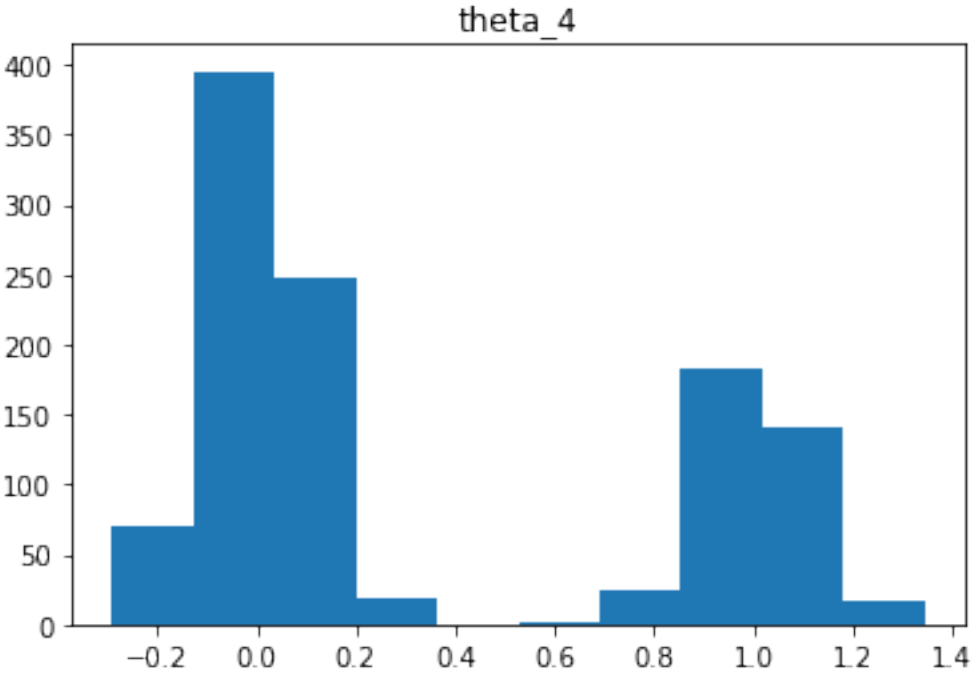
[ ]:
```
for i in df.columns:
    plt.hist(df[i])
```

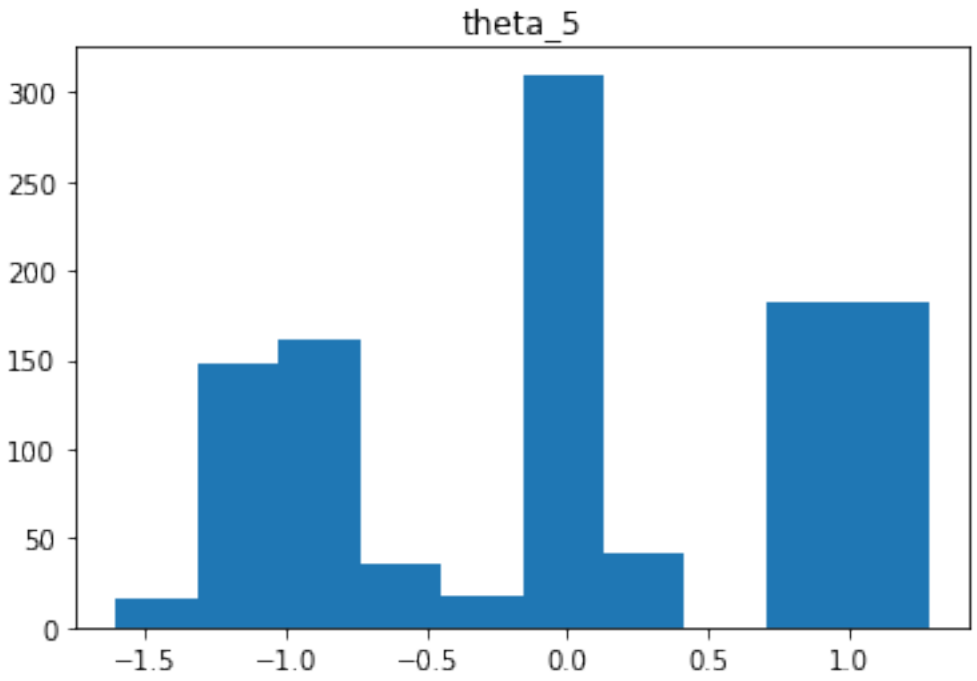
```
    plt.title(i)
    plt.show()
```

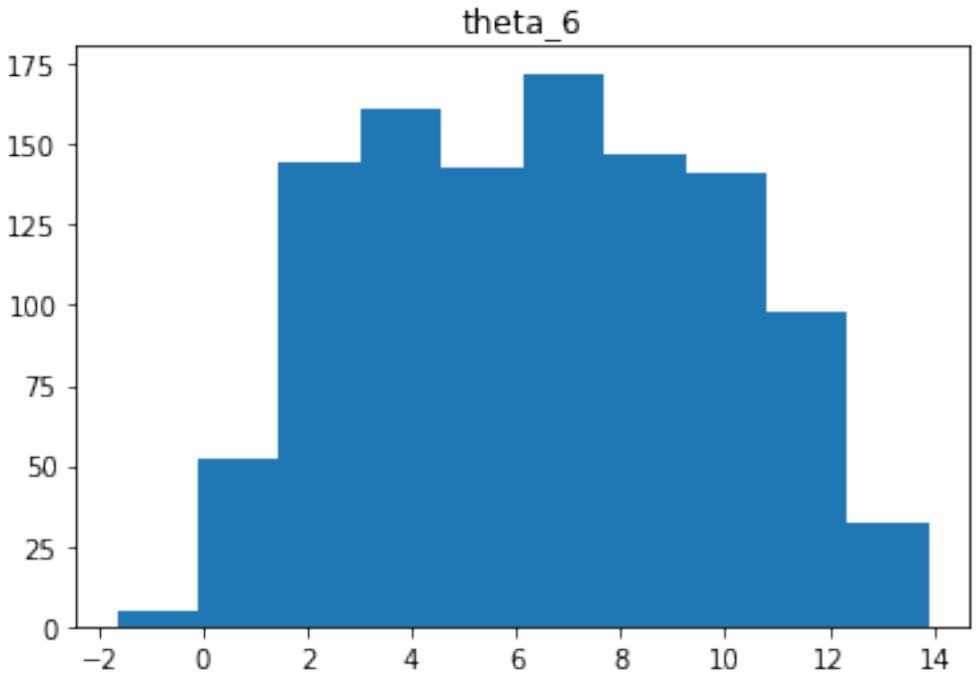





theta_3
500
400
300
200
100
0
−10
−8
−6
−4
−2
0
2
4
6


theta_4
400
350
300
250
200
150
100
50
0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

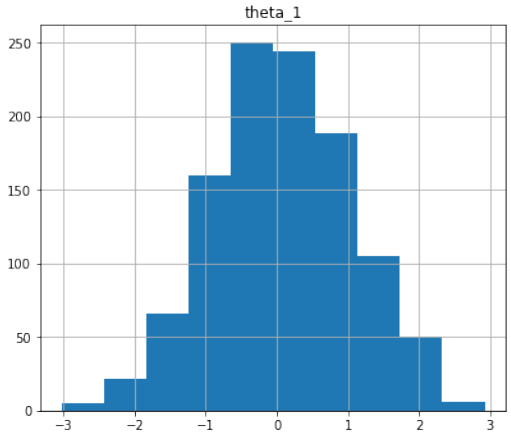


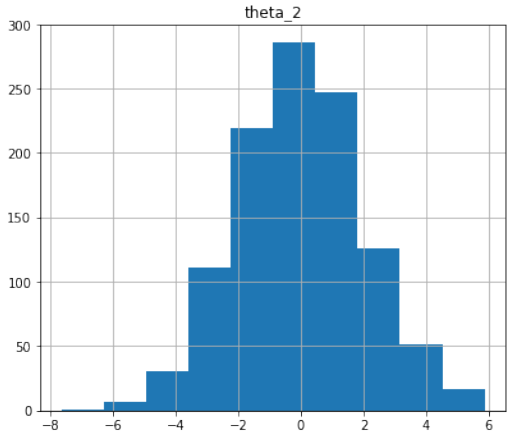


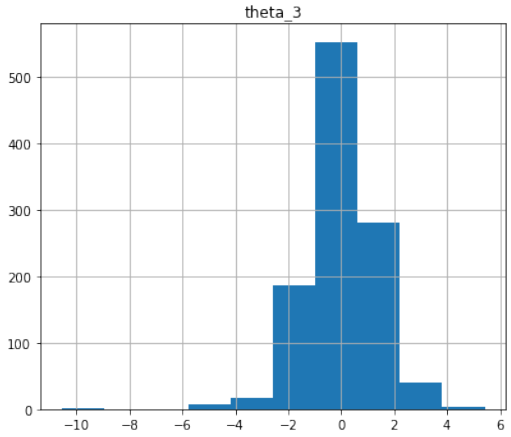
[ ]:
```
fig = plt.figure(figsize = (15,20))
ax = fig.gca()
```

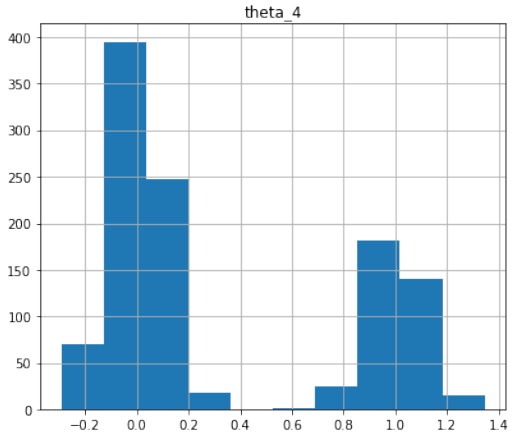
```
df.hist(ax = ax)
plt.show()
```

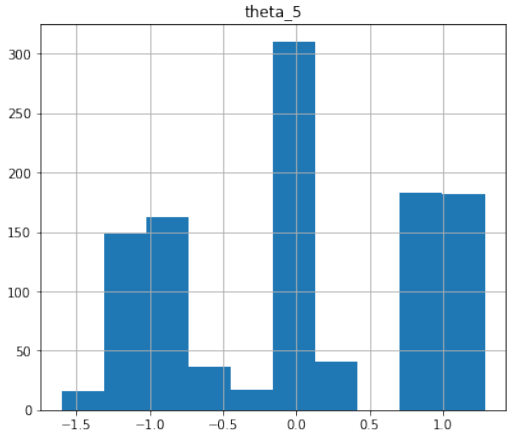
```
C:\Users\norbe\AppData\Local\Temp\ipykernel_12604\111277950.py:3: UserWarning:
To output multiple subplots, the figure containing the passed axes is being
cleared.
  df.hist(ax = ax)
```

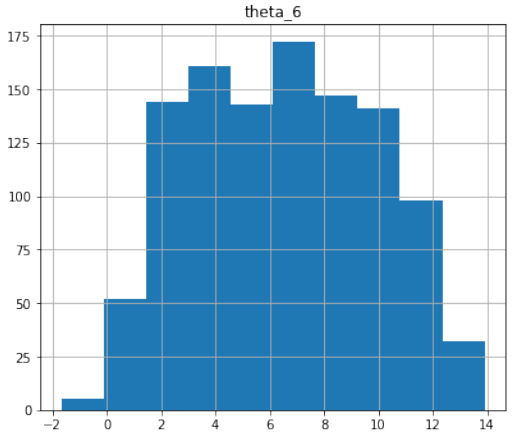




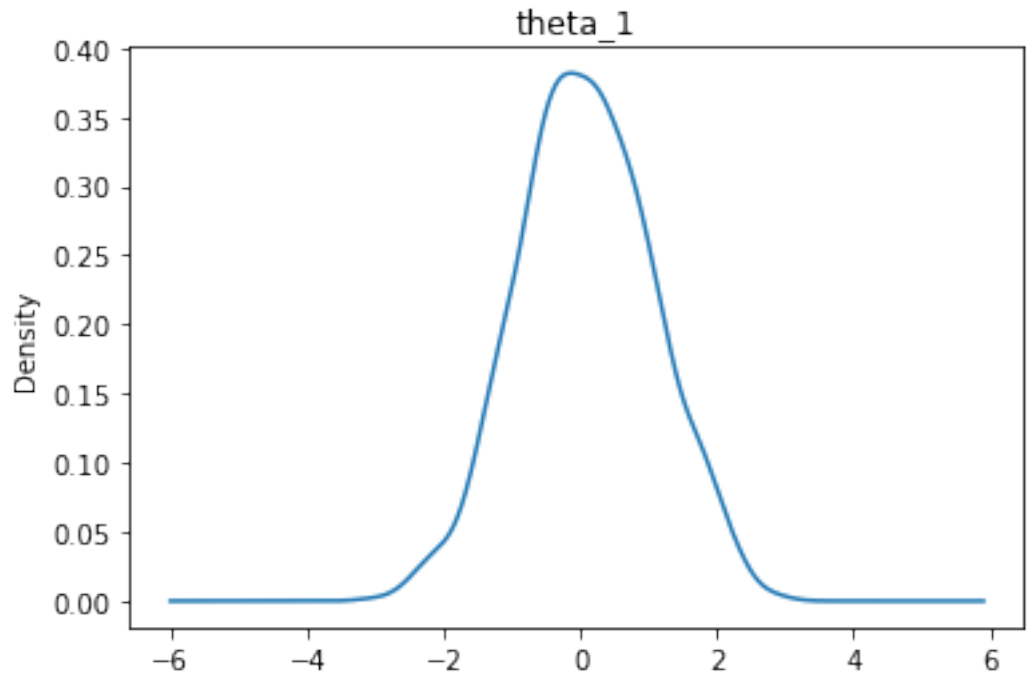


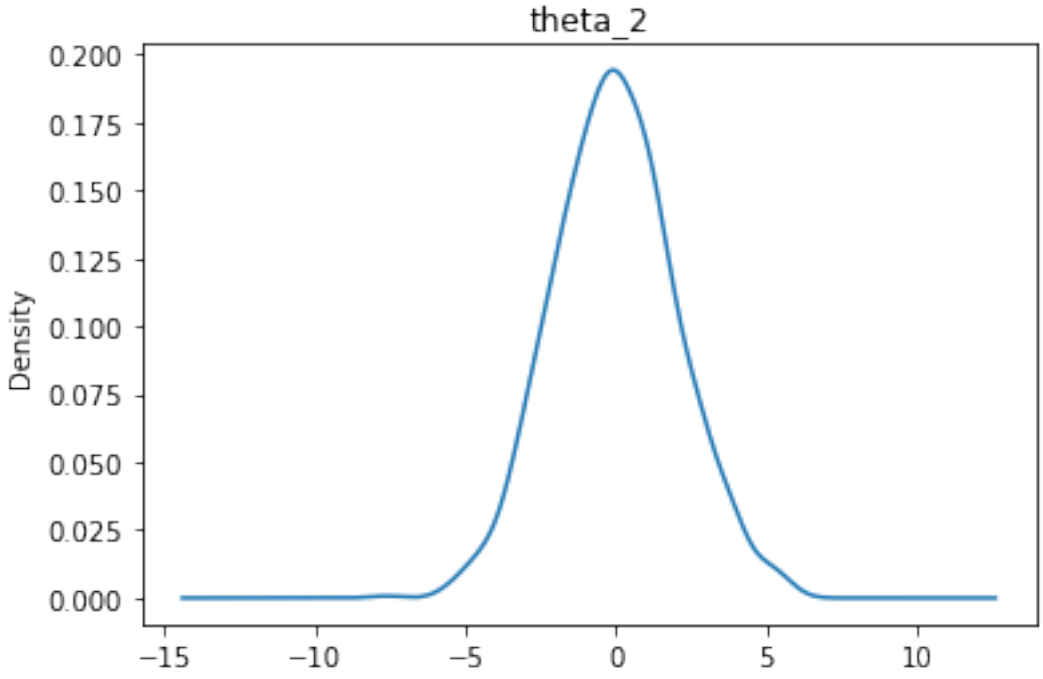


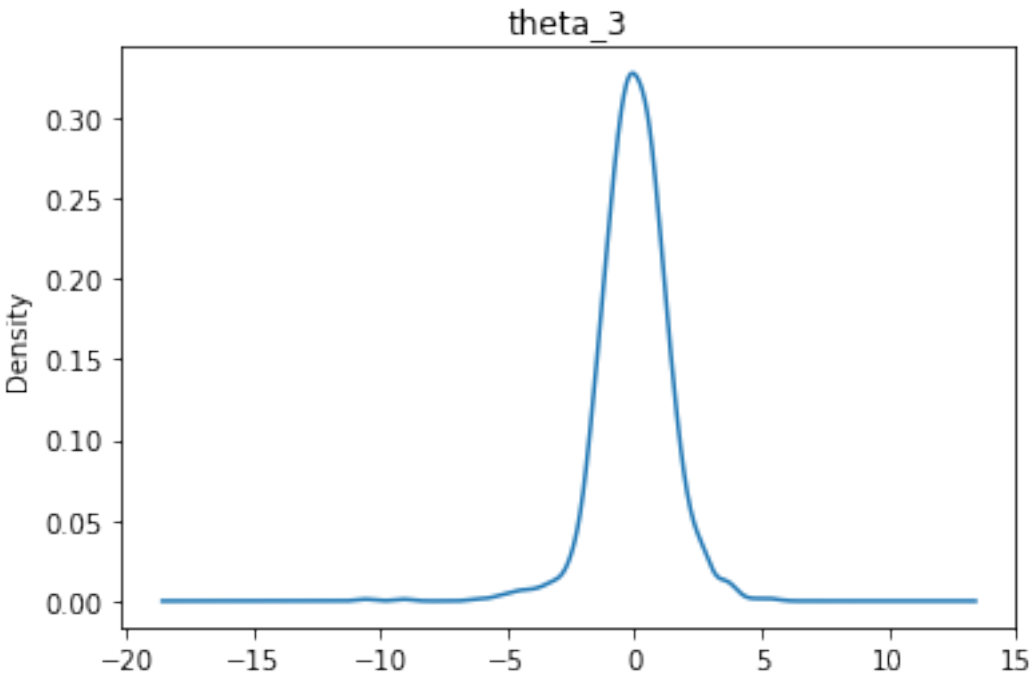


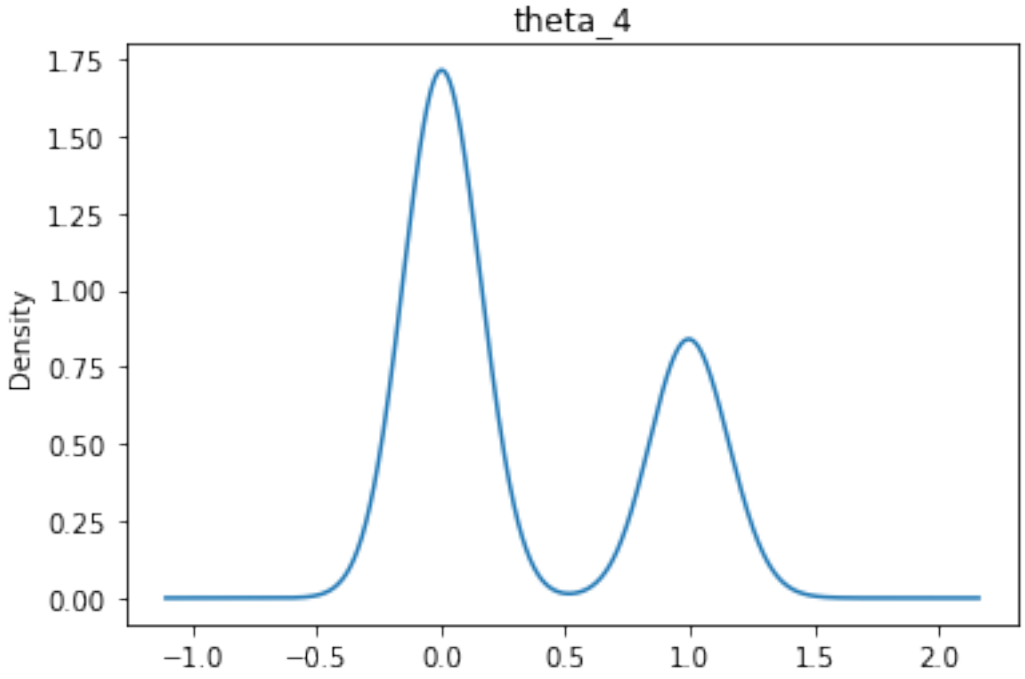


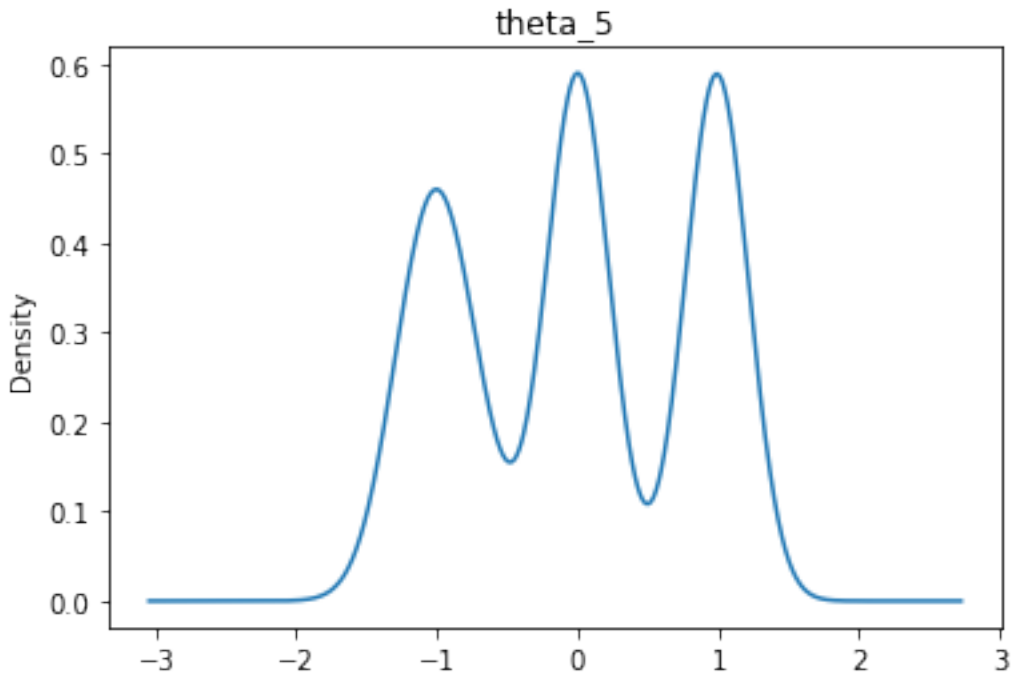
[ ]:
```
for i in df.columns:
    ax = df[i].plot.kde()
    plt.title(i)
    plt.show()
```

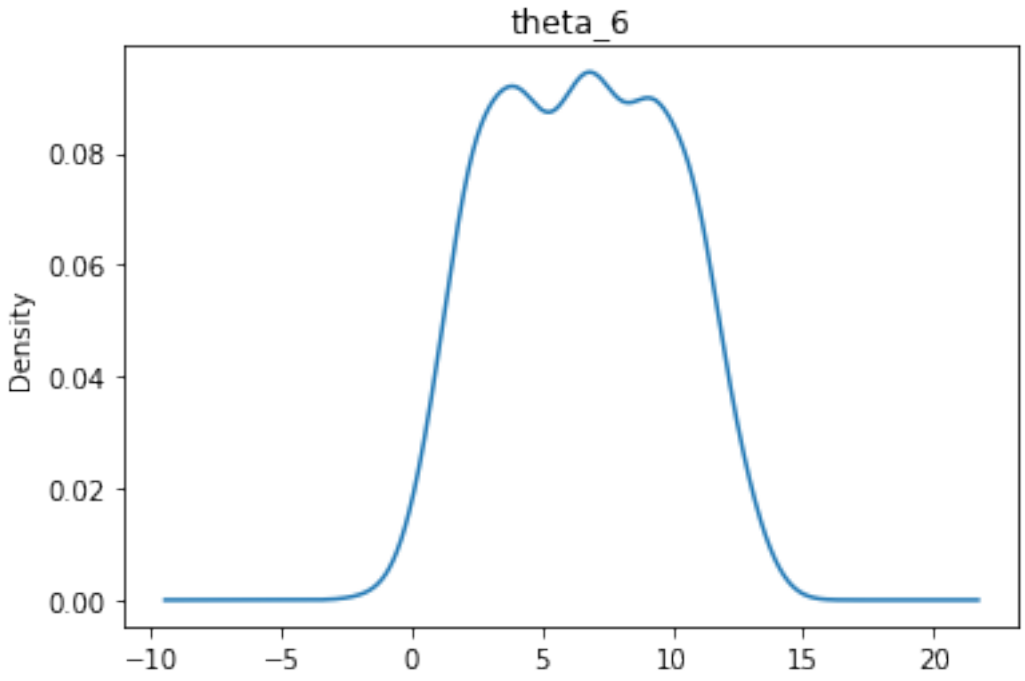

theta_1
Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−6
−4
−2
0
2
4
6


theta_2
Density
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
−15
−10
−5
0
5
10


theta_3
Density
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−20
−15
−10
−5
0
5
10
15


theta_4
Density
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0





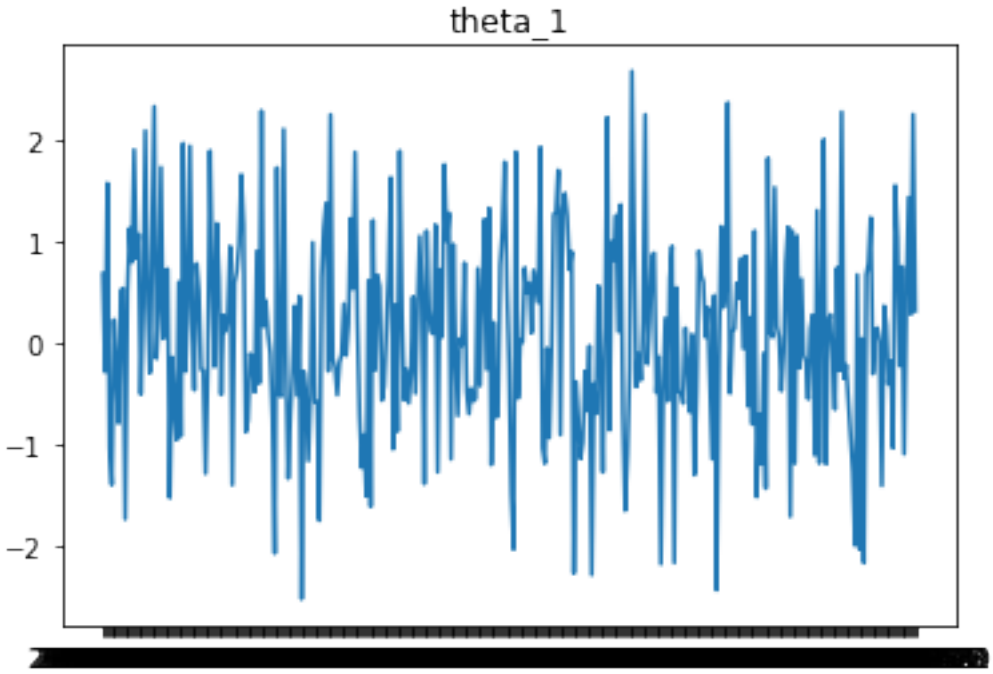
[ ]:
```
df_2018 = df.loc['2018-01-01':'2018-12-31'][["theta_1", "theta_2", "theta_3",␣
↪"theta_4"]]
```

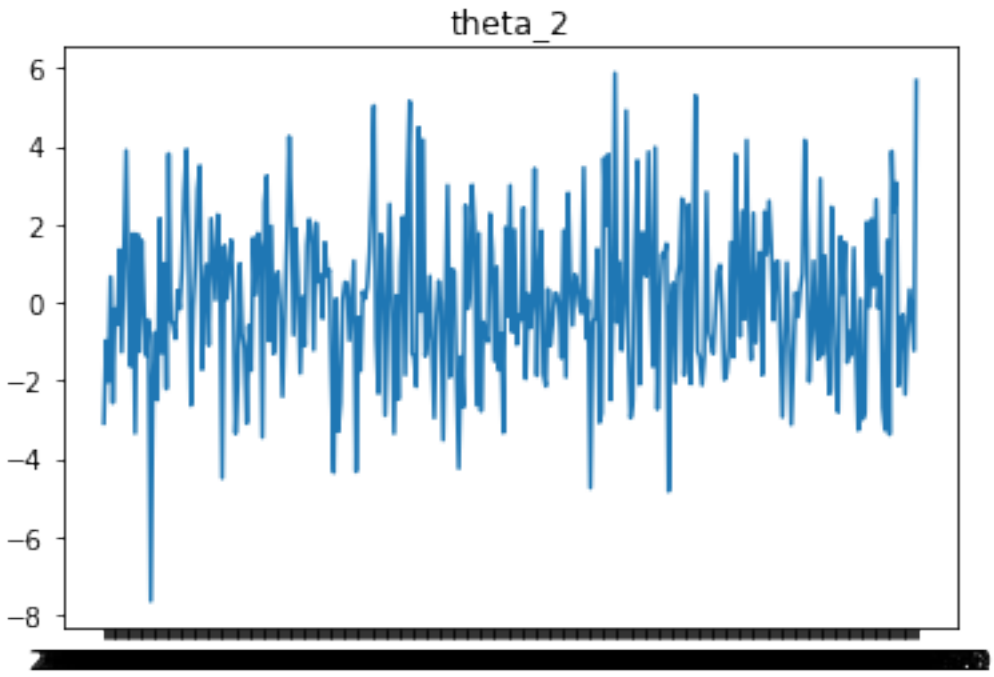
```
df_2018.head()
```

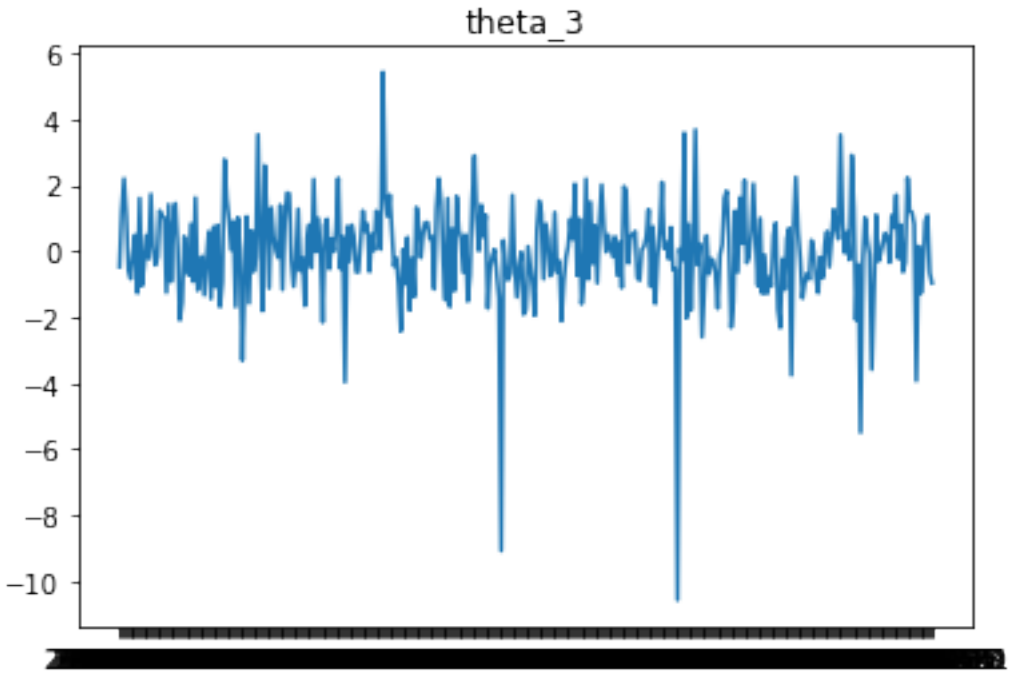
[ ]:

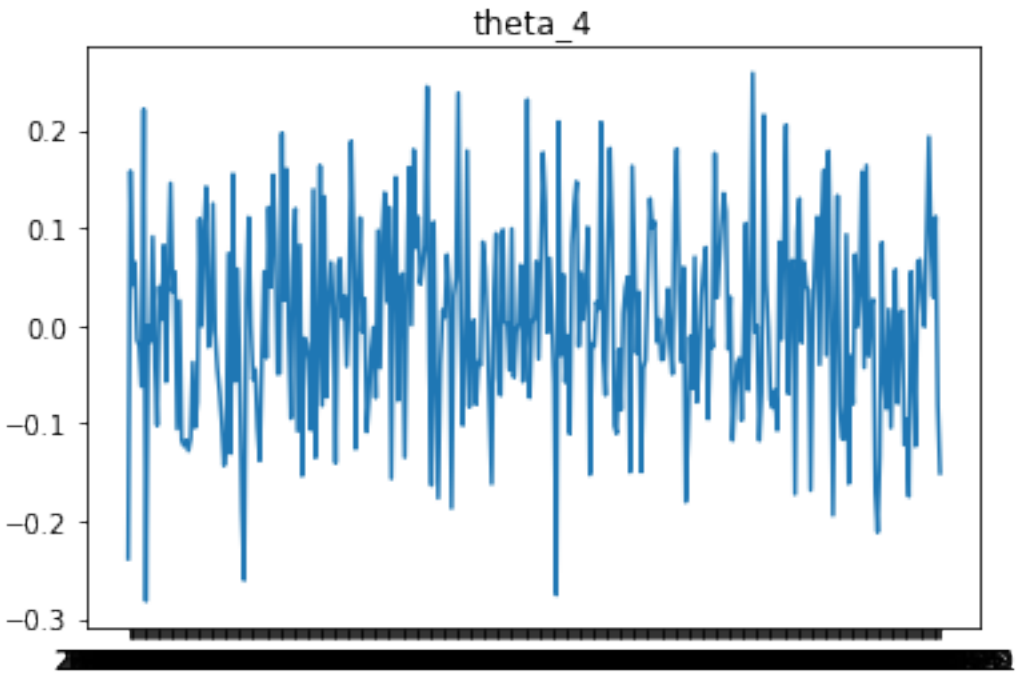
	theta_1	theta_2	theta_3	theta_4
2018-01-01	0.682693	-3.091767	-0.475717	-0.238530
2018-01-02	-0.283107	-0.979955	1.233933	0.158031
2018-01-03	1.572221	-2.033528	2.196317	0.041347
2018-01-04	-1.042981	0.651530	1.060125	0.064832
2018-01-05	-1.392614	-2.570905	-0.600063	-0.015025

[ ]:
```
for i in df_2018.columns:
    plt.plot(df_2018.index, df_2018[i])
    plt.title(i)
    plt.show()
```

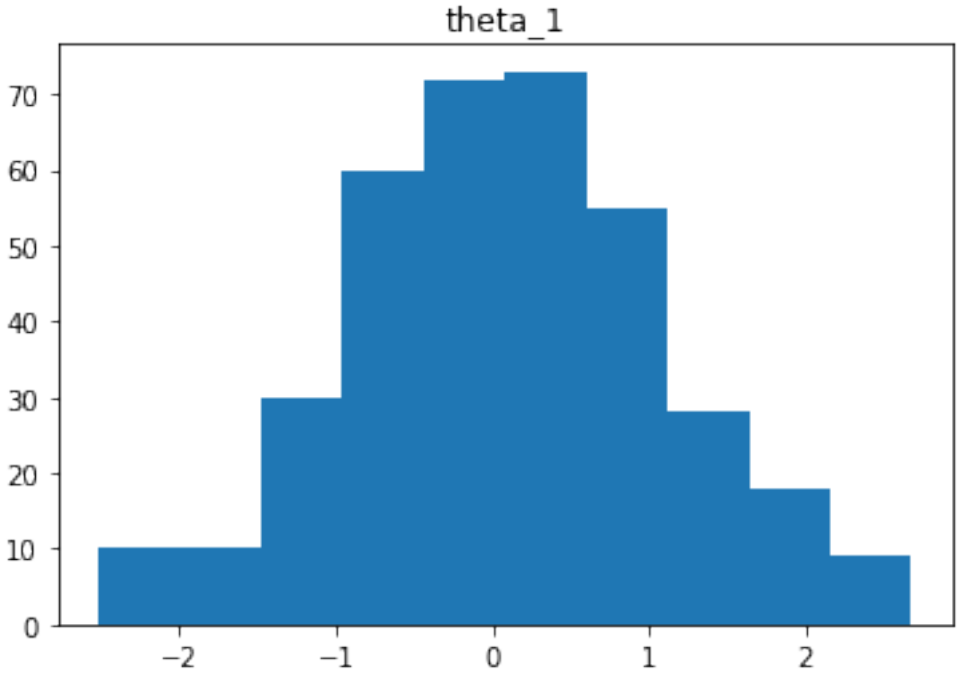


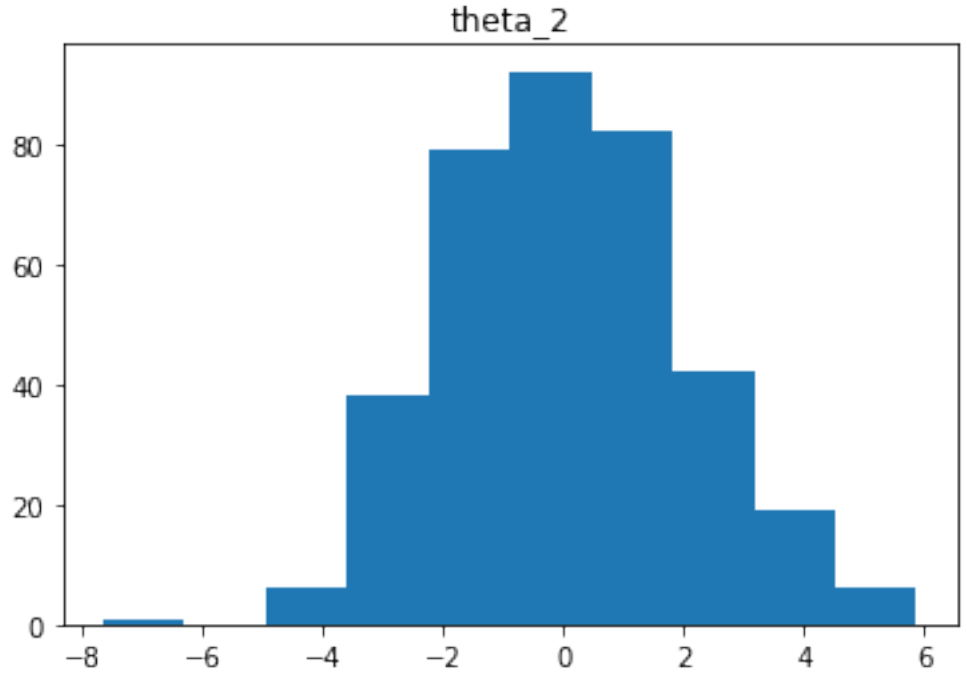

theta_2
6
4
2
0
−2
−4
−6
−8

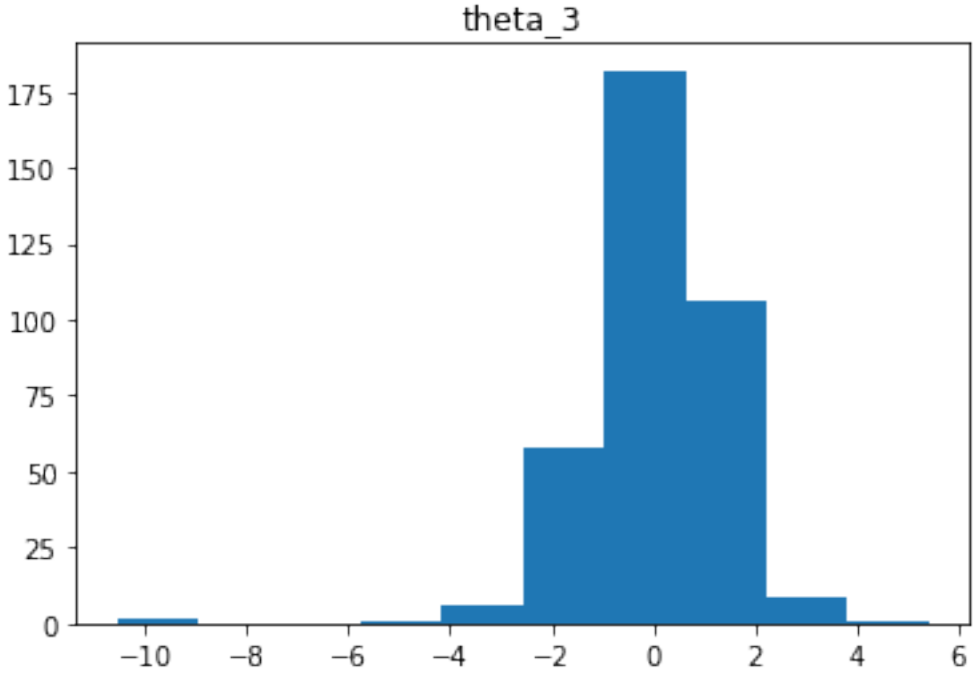

theta_3
6
4
2
0
−2
−4
−6
−8
−10

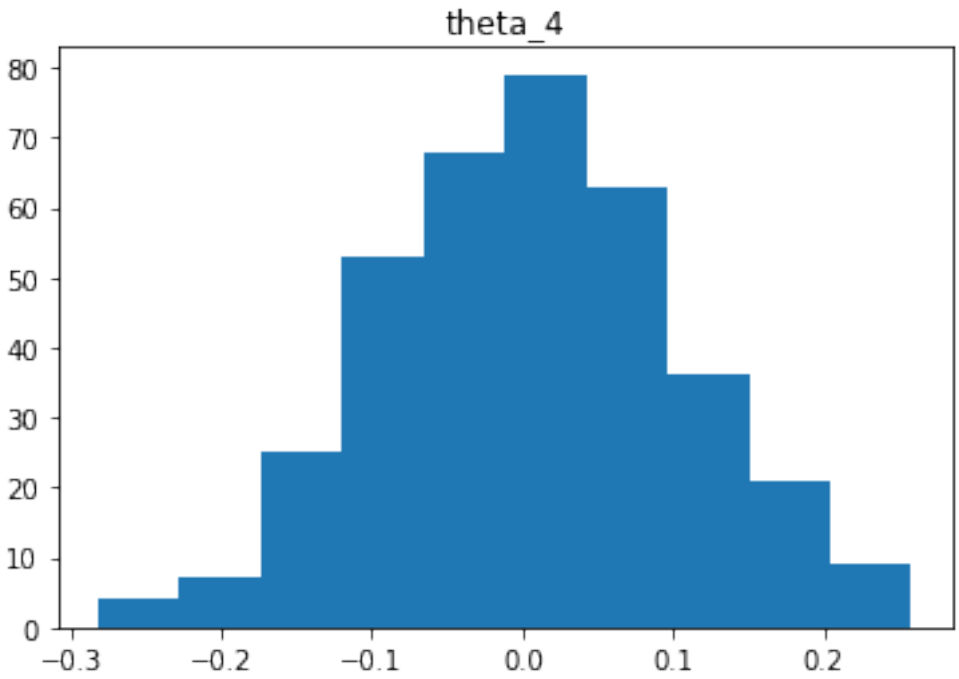


[ ]:
```
for i in df_2018.columns:
    plt.hist(df_2018[i])
    plt.title(i)
    plt.show()
```



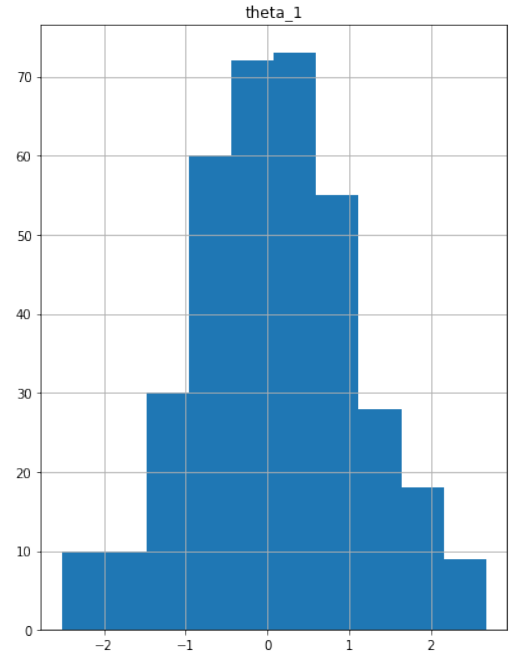

theta_2
80
60
40
20
0
−8
−6
−4
−2
0
2
4
6

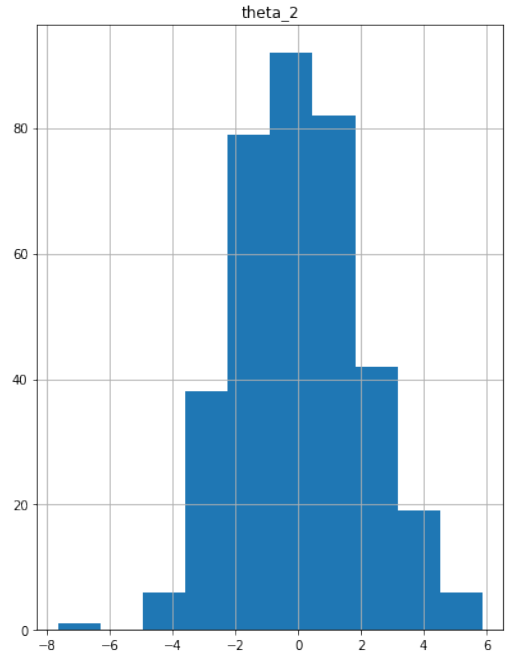

theta_3
175
150
125
100
75
50
25
0
−10
−8
−6
−4
−2
0
2
4
6

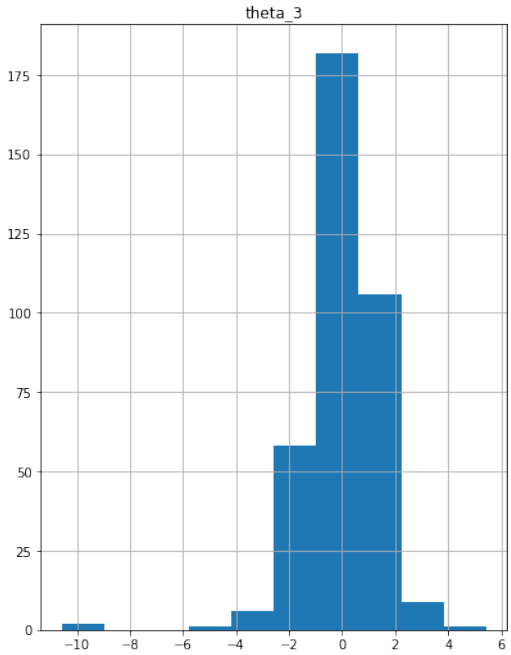


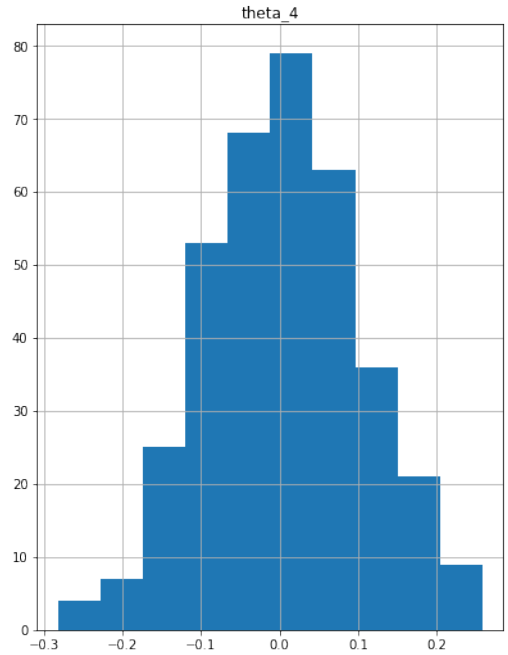
[ ]:
```
fig = plt.figure(figsize = (15,20))
ax = fig.gca()
df_2018.hist(ax = ax)
plt.show()
```

```
C:\Users\norbe\AppData\Local\Temp\ipykernel_12604\712917149.py:3: UserWarning:
To output multiple subplots, the figure containing the passed axes is being
cleared.
  df_2018.hist(ax = ax)
```

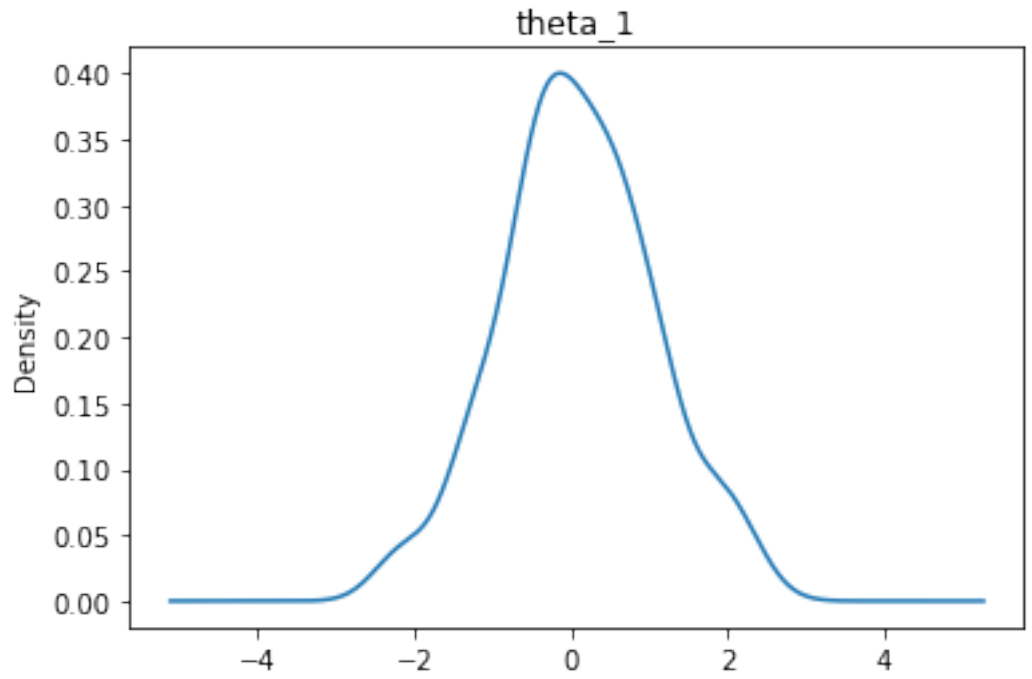


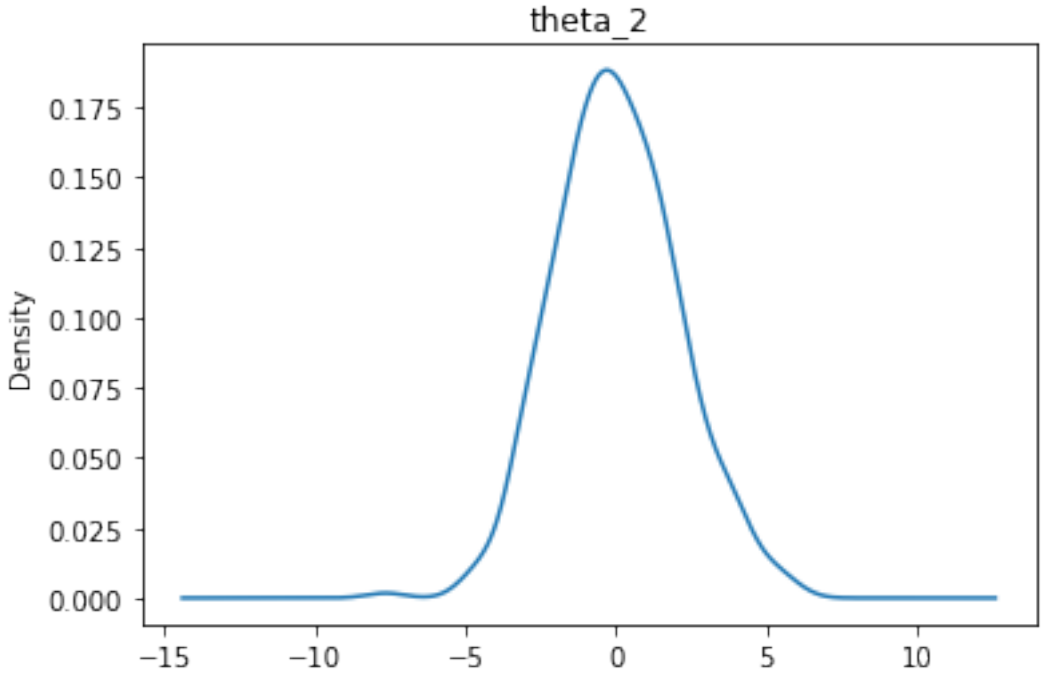


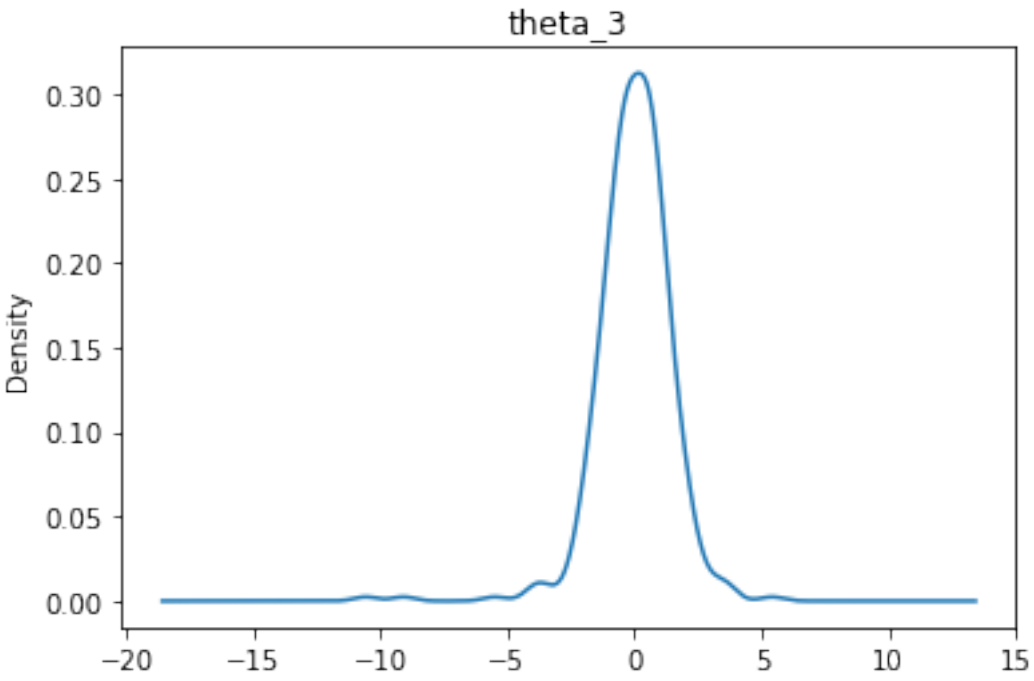


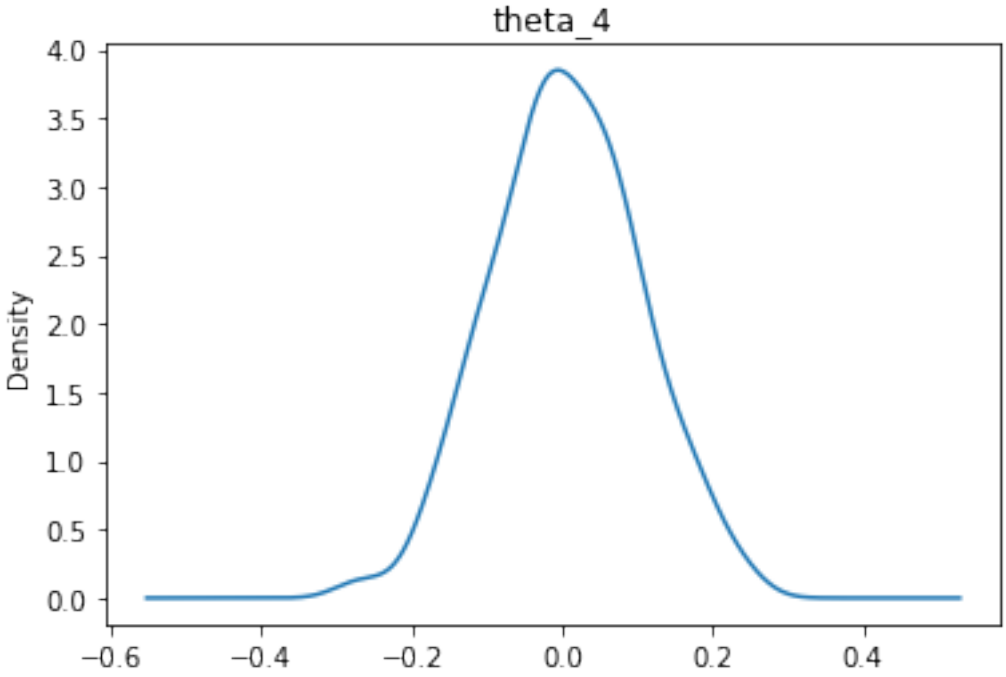


[ ]:
```
for i in df_2018.columns:
    ax = df_2018[i].plot.kde()
    plt.title(i)
    plt.show()
```


theta_1
Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−4
−2
0
2
4


theta_2
Density
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
−15
−10
−5
0
5
10


theta_3
Density
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−20
−15
−10
−5
0
5
10
15


theta_4
Density
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.6
−0.4
−0.2
0.0
0.2
0.4

1.2 Excersise 2

[ ]:
```
from cmdstanpy import CmdStanModel
```

[ ]:
```
# F=7 L=3

data = {"N": 10, "y": [0, 1, 0, 0, 1, 1, 0, 0, 0, 0]}

model = CmdStanModel(stan_file="bern_1.stan")
```

```
INFO:cmdstanpy:found newer exe file, not recompiling
```

[ ]:
```
fit = model.sample(data=data, output_dir='out')
theta = fit.stan_variable("theta")
```

```
INFO:cmdstanpy:CmdStan start processing
chain 1 |          | 00:00 Status


chain 1 |       | 00:00 Iteration: 1100 / 2000 [ 55%]  (Sampling)


chain 1 |      | 00:00 Sampling completed
chain 2 |      | 00:00 Sampling completed
chain 3 |      | 00:00 Sampling completed
chain 4 |      | 00:00 Sampling completed



INFO:cmdstanpy:CmdStan done processing.
```

[ ]:
```
df = fit.summary()

df
```

[ ]:

name	Mean	MCSE	StdDev	5%	50%	95%	N_Eff	N_Eff/s	R_hat
lp__	-8.20	0.0200	0.72	-9.60	-7.90	-7.60	1300.0	6300.0	1.0
theta	0.33	0.0033	0.13	0.14	0.32	0.57	1600.0	8100.0	1.0

[ ]:
```
df_theta = df.loc['theta']

mean = theta.mean()
median = df_theta["50%"]
q5 = df_theta["5%"]
q95 = df_theta["95%"]
```

[ ]:
```
plt.hist(theta, bins=50, density=True)
plt.axvline(mean, color='y')
plt.axvline(median, color='b')
plt.axvline(q5, color='r')
plt.axvline(q95, color='g')
plt.show()
```