Fintech 545

Week 2 Project

1/26/2024

Yi Wang

Problem 1

a. Using normalized formula, the four moments are (mean, variance, skewness, and "excess" kurtosis): 1.0489703904839585, 5.4217934611998455, 0.8806086425277359, 23.122200789989723.

b. Using my chosen package, scipy.stats, the four moments are (mean, variance, skewness, and "excess" kurtosis): 1.0489703904839585, 5.4217934611998455, 0.8806086425277364, 23.122200789989723.

c. To determine whether my statistical package functions are biased, I created a normally distributed sample first. Then, I collected kurtosis value for each sample and calculated the t-statistics for testing whether the average kurtosis value differs from 3. Using the result of t-statistics, I got p-values for the t-test. After that, I calculated the p-value for the t-test using ttest_isamp from the package scipy.stats I used for 1b. Comparing the p-values, I found they are not very close to each other, indicating my statistical package functions are biased.

```
0.0 0.006672117295280352
Match the stats package test?: False
```

Problem 2

a. Result fitting data using OLS:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.346
Model:                            OLS   Adj. R-squared:                  0.342
Method:                 Least Squares   F-statistic:                     104.6
Date:                Fri, 26 Jan 2024   Prob (F-statistic):           5.59e-20
Time:                        19:33:27   Log-Likelihood:                -284.54
No. Observations:                 200   AIC:                             573.1
Df Residuals:                     198   BIC:                             579.7
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
one           -0.0874      0.071     -1.222      0.223      -0.228       0.054
x              0.7753      0.076     10.226      0.000       0.626       0.925
==============================================================================
Omnibus:                       11.922   Durbin-Watson:                   2.023
Prob(Omnibus):                  0.003   Jarque-Bera (JB):               16.685
Skew:                           0.387   Prob(JB):                     0.000238
Kurtosis:                       4.184   Cond. No.                         1.09
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
one   -0.087384
x      0.775274
dtype: float64 one    0.071496
x      0.075814
dtype: float64 1.006275159813318
```

Result fitting data using MLE given the assumption of normality:

```
  message: Optimization terminated successfully.
  success: True
   status: 0
      fun: 284.53756305442784
        x: [-8.738e-02  7.753e-01  1.004e+00]
      nit: 14
      jac: [-3.815e-06  0.000e+00  3.815e-06]
 hess_inv: [[ 5.145e-03 -4.960e-04  5.825e-06]
            [-4.960e-04  5.893e-03  6.009e-06]
            [ 5.825e-06  6.009e-06  2.591e-03]]
     nfev: 72
     njev: 18
[-0.08738449  0.77527409  1.00375631]
```

According to the results, I found that the fitted beta and standard deviation of the OLS error is very close to the fitted beta and standard deviation of the MLE error.

b. Result fitting data using MLE given the assumption of T distribution of errors:

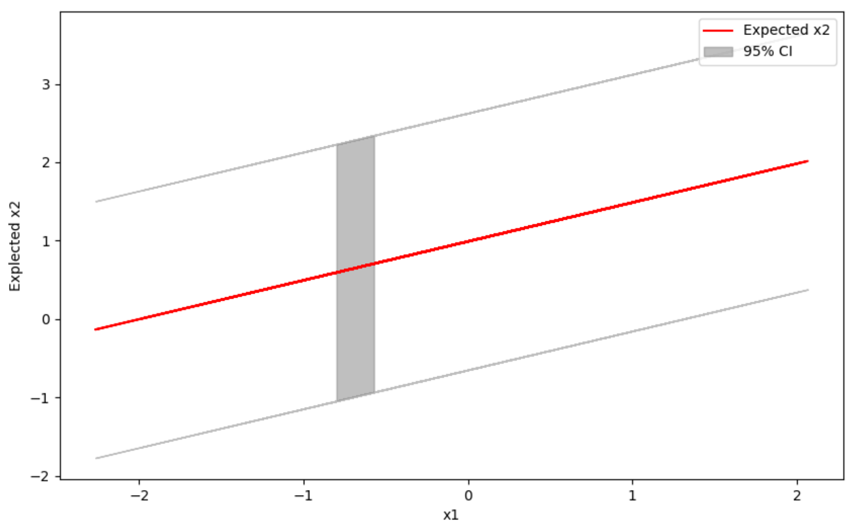
```
   message: Optimization terminated successfully.
   success: True
    status: 0
       fun: 281.29340317987743
         x: [ 8.551e-01  7.160e+00 -9.727e-02  6.750e-01]
       nit: 18
       jac: [ 3.815e-06  3.815e-06  3.815e-06  0.000e+00]
  hess_inv: [[ 5.096e-03  1.672e-01 -3.321e-05  2.251e-03]
             [ 1.672e-01  1.165e+01 -3.701e-03  1.288e-01]
             [-3.321e-05 -3.701e-03  4.825e-03 -4.434e-04]
             [ 2.251e-03  1.288e-01 -4.434e-04  6.921e-03]]
      nfev: 110
      njev: 22
[ 0.85510487  7.15984961 -0.09726944  0.67500978]
AIC_norm < AIC_t: False
BIC_norm < BIC_t: False
```

The fitted beta and standard deviation of this model are all a little bit smaller than the previous one. Using AIC and BIC, both values of the MLE (normality) are greater than the MLE (T), indicating MLE given the assumption of T distribution of errors is a better fit for the given dataset.

c. Results fitting data using MLE given X = [X1, X2]:

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      x   No. Observations:                  200
Model:                 ARIMA(0, 0, 3)   Log Likelihood                -269.585
Date:                Fri, 26 Jan 2024   AIC                            549.169
Time:                        19:33:28   BIC                            565.661
Sample:                             0   HQIC                           555.843
                                - 200
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0622      0.079      0.787      0.431      -0.093       0.217
ma.L1          0.0503      0.073      0.692      0.489      -0.092       0.193
ma.L2          0.0225      0.075      0.298      0.766      -0.125       0.170
ma.L3          0.1226      0.076      1.620      0.105      -0.026       0.271
sigma2         0.8674      0.098      8.867      0.000       0.676       1.059
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                 1.47
Prob(Q):                              0.92   Prob(JB):                         0.48
Heteroskedasticity (H):               0.89   Skew:                             0.02
Prob(H) (two-sided):                  0.65   Kurtosis:                         2.58
===================================================================================
```

Distribution of X2:



Problem 3

I applied ARIMA, which is from statsmodels.tsa.arima.model, to build the six models. After that, I used model.aic to get their aic values (from AR1 to MA3):

```
1644.6555047688475
1581.0792659049775
1436.6598066945876
1567.403626370787
1537.9412063807388
1536.867708735031
```

AR3 has the smallest value, indicating it is the best fit.

I also plotted these models to double check my result: