Week 2 Project

Question 1

For conditional distribution of the Multivariate Normal, we first get the partition X, μ, and Σ:

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} -0.0223 \\ -0.1405 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 0.8989 & 0.5629 \\ 0.5629 & 1.3152 \end{bmatrix}$$

Therefore, given $x_2 = a$, then $x_1 \sim N(\bar{\mu}, \overline{\Sigma})$, where $\bar{\mu} = \mu_1 + \Sigma_{12}{\Sigma_{22}}^{-1}(a - \mu_2) = 0.4280 * a + 0.0379$.

Then we run the OLS estimation, and get the result $\hat{y} = 0.4280 * x + 0.0379$

```
ols = sm.OLS(x1, x2).fit()
print(ols.summary())
```

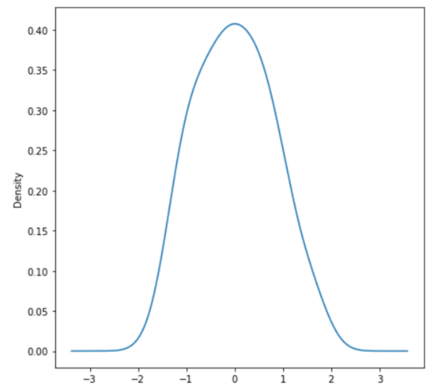
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.268
Model:                            OLS   Adj. R-squared:                  0.261
Method:                 Least Squares   F-statistic:                     35.89
Date:                Thu, 13 Jan 2022   Prob (F-statistic):           3.47e-08
Time:                        17:51:31   Log-Likelihood:                -120.46
No. Observations:                 100   AIC:                             244.9
Df Residuals:                      98   BIC:                             250.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0379      0.082      0.461      0.646      -0.125       0.201
x              0.4280      0.071      5.990      0.000       0.286       0.570
==============================================================================
Omnibus:                        5.101   Durbin-Watson:                   2.006
Prob(Omnibus):                  0.078   Jarque-Bera (JB):                2.716
Skew:                           0.145   Prob(JB):                        0.257
Kurtosis:                       2.246   Cond. No.                         1.20
==============================================================================
```

We can see that the coefficients are the same between two methods. The values are same because the linear regression is based on the conditional distribution, and the predicted value of y is calculated with a given x.

Question 2

Running OLS regression, we get the result $\hat{y} = 0.4280 * x + 0.0379$, and the distribution of the error term is similar to the normal distribution.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.268
Model:                            OLS   Adj. R-squared:                  0.261
Method:                 Least Squares   F-statistic:                     35.89
Date:                Thu, 13 Jan 2022   Prob (F-statistic):           3.47e-08
Time:                        20:41:06   Log-Likelihood:                -120.46
No. Observations:                 100   AIC:                             244.9
Df Residuals:                      98   BIC:                             250.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0379      0.082      0.461      0.646      -0.125       0.201
x              0.4280      0.071      5.990      0.000       0.286       0.570
==============================================================================
Omnibus:                        5.101   Durbin-Watson:                   2.006
Prob(Omnibus):                  0.078   Jarque-Bera (JB):                2.716
Skew:                           0.145   Prob(JB):                        0.257
Kurtosis:                       2.246   Cond. No.                         1.20
==============================================================================
```



Then we define the MLE function and fit the data with MLE given the assumption of normality.

```
      fun: 120.46049859542816
 hess_inv: <3x3 LbfgsInvHessProduct with dtype=float64>
      jac: array([ 3.12638804e-05, -3.12638804e-05,  9.94759825e-06])
  message: 'CONVERGENCE: REL_REDUCTION_OF_F_<=_FACTR*EPSMCH'
     nfev: 56
      nit: 8
     njev: 14
   status: 0
  success: True
        x: array([0.03787704, 0.42800357, 0.80707917])
```

The result is same as the OLS result, which means that MLE fits the assumption of normality.

Then we create a random T distribution with 99 degree of freedom and 100 samples, and we fit it with MLE method.

```
t.fit(t.rvs(df=99, size=100, loc=0, scale=1))
```

```
(1721939.2535913307, 0.042192545269869444, 1.0139618964399268)
```

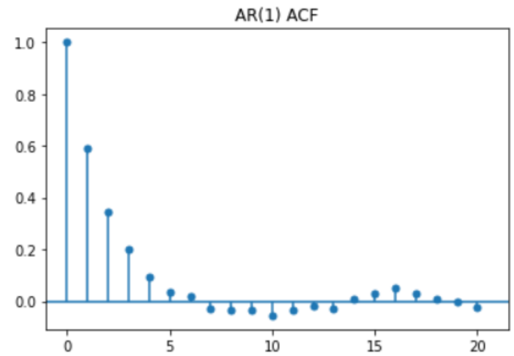
The result shows that MLE does not fit in the assumption of a T distribution of the errors.

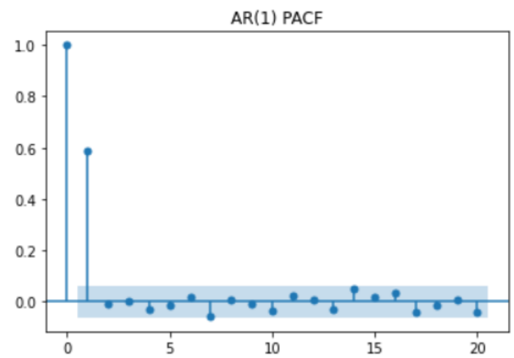
Then we compare the moments of the regression with the normal distribution, we get mean = -0.141, standard deviation = 1.141, skewness = 0.145, kurtosis = 2.246, which are close to the standard normal.

If the normality assumption breaks, $E[y_i] = \beta_0 + \beta_1 x_i + \varepsilon_i$ will be biased because $E[\varepsilon_i] \neq 0$, which will make the estimation invalid.

Question 3

When doing estimation using time series, we need to consider ACF and PACF together. For AR process, we expect ACF decrease gradually while the PACF should have a sharp drop after p significant lags. On the other hand, MA has the opposite features with ACF has a sharp drop a certain q number of lags while PACF decrease gradually.





We can distinguish AR and MA by looking at the place where a sharp drop happens and find the appropriate model by counting how many time-lag pass when the sharp drop happens. (TowardDataScience, 2020)

References

ThoughtPartnerforData (2020, August 13). *Identifying AR and Ma terms using ACF and PACF plots in time series forecasting*. Medium. Retrieved January 14, 2022, from https://towardsdatascience.com/identifying-ar-and-ma-terms-using-acf-and-pacf-plots-in-time-series-forecasting-ccb9fd073db8