Assignment 2 Solution

Md Mohidul Haque
Immatriculation number: 619563

1.
I have imported two data sets (by using pandas) as the names df_unemploymentRate (unemployment rate) and df_unfilled_job (unfilled job vacancies). Both data sets have two rows, and I renamed the column names. I merged those two data sets in the 'inner' way and stored them in df. Then sorted all the values of 'df' regarding the 'Date' column. After merging data sets, df_unemploymentRate data loses 424 rows, and df_unfilled_job data lose 36 rows. In that case, we have a lot of missing data, which is not perfect for analysis. Even if we start our statistical comparison from 1969-01-01, we are still missing data in some places.

2.
After calculating the growth rates (where periods are 4) of the Unemployment rate as Growth_UR and unfilled job vacancies as Growth_UJ, we found that the correlation of Growth_UR and Growth_UJ is Negative (-0.648505).In the Scatter plot, we saw that kind of relationship.

Further, in the regression, we use Growth_UJ as a dependent variable. We found that the coefficient of Growth_UR (-0.5736) and the confidence interval of Growth_UR are also negative ([-0.666, -0.481]).
The standard error of Growth_UR is larger than it's coefficient. And t statistics value for Growth_UR (-12.198) are smaller than it's coefficient (-0.5736). P-value looks good, reject the null hypothesis. So, unemployment rate (independent variable) is negatively correlated with the dependent variable unfilled job vacancies.

There is very little difference between the values of R-squared (0.421) and Adj. R-squared (0.418). F-statistic should be large (large F-statistic means the model has good impact), and Prob (F-statistic) should be small. In this case, F-statistic is 148.8, model is quite sufficient.

Omnibus (0.539) is small, and Prob(Omnibus) is 0.764, which is close to one, so it can be said that the residuals follow Gaussian distribution. And

Durbin-Watson (0.231) values are not between 1 and 2. Prob (JB) is 0.717, which is close to one, hence residuals are normally distributed. Skew (0.041) is near to zero, means normal distribution. By considering those, we can say that, the relation is linear.

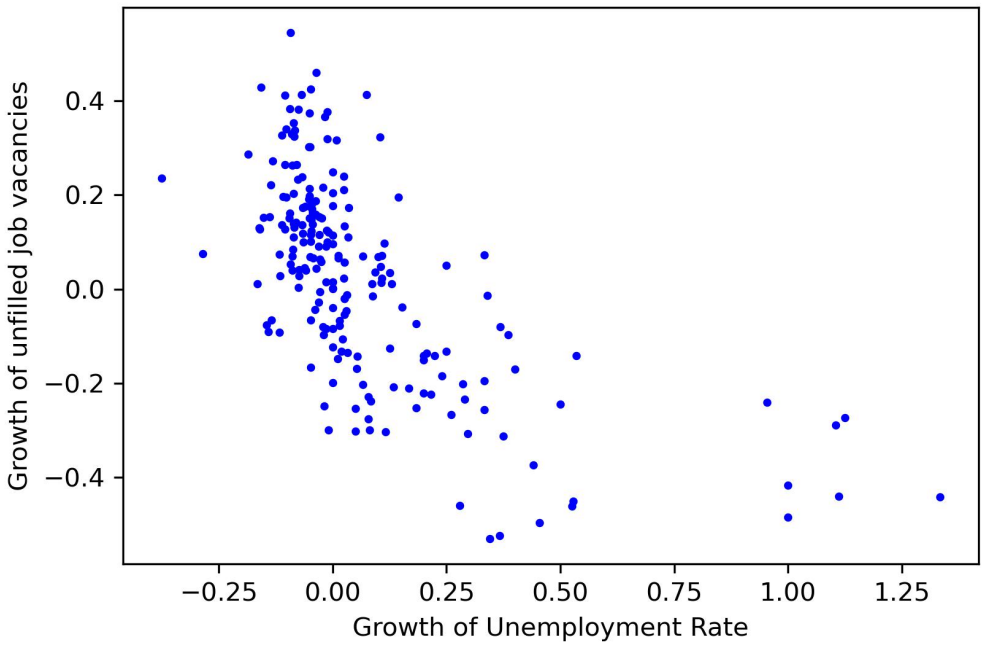
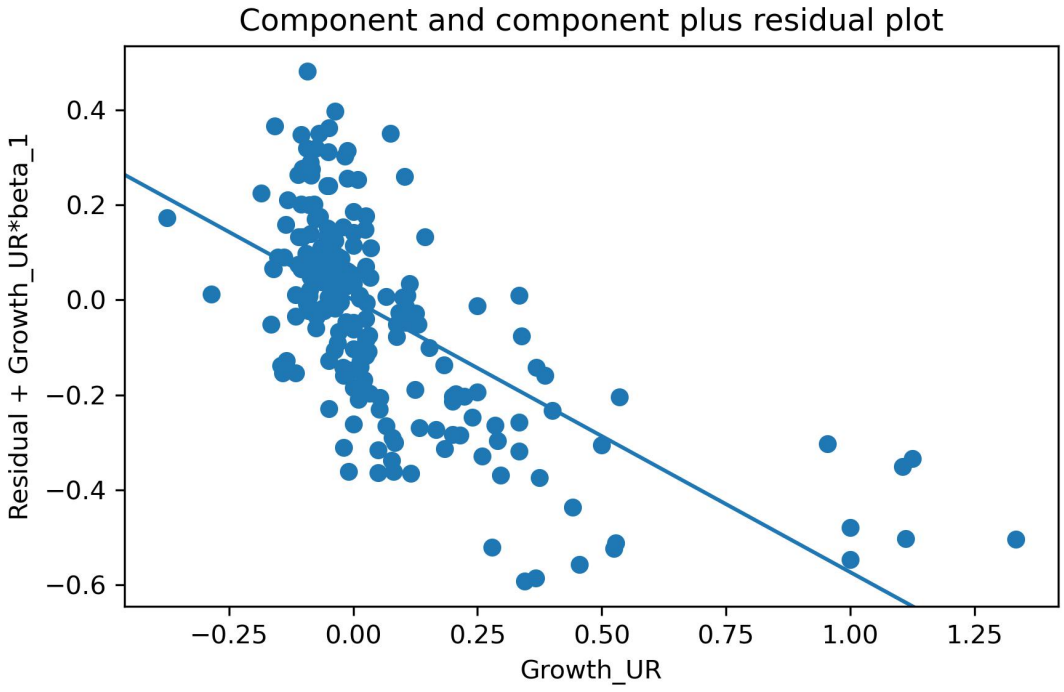
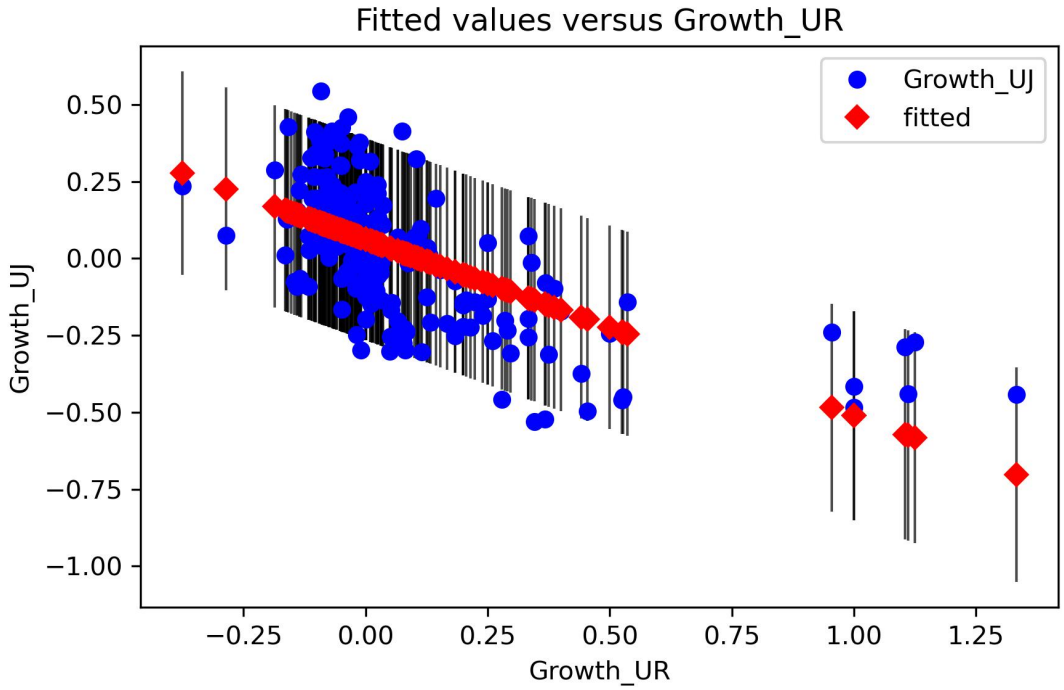


Fig: Scatter plot of the Growth of Unfilled job vacancies against unemployment rate.

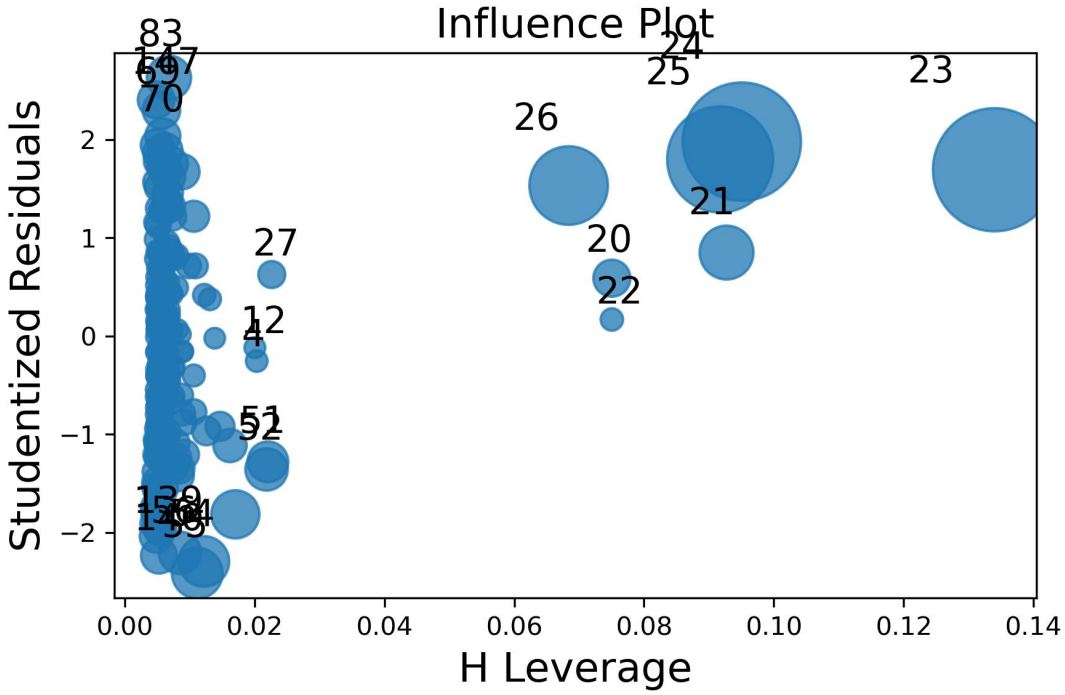
From the CCPR plot , we see that there is negative correlation between Unemployment rate and Unfilled job vacancies.



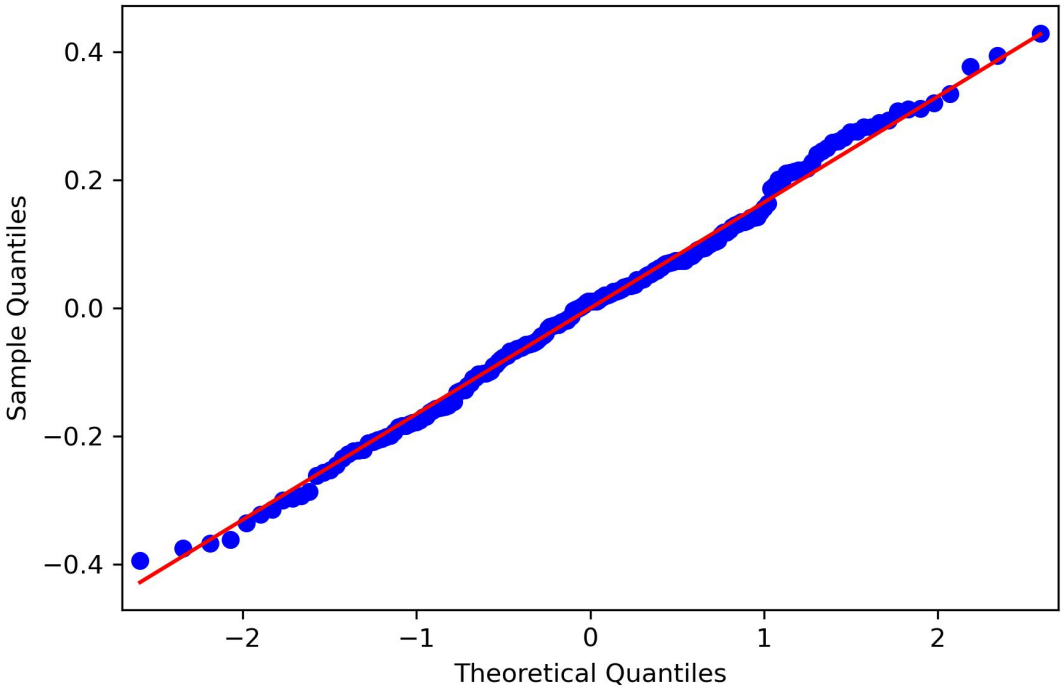
In the fit plot, most of the plotted data are in the range of confidence interval.



In the influence plot, observation 20 to 26 have some high influence.



From the QQ plot, the residuals are normally distributed, and the relationship is linear.



In the histogram plot, most of the dominating frequency for both growth rates are in between the interval -0.25 to 0. It's not totally normally distributed.