Estimate a regression model using the given data.

a. Plot the residuals obtained and see if you observe any systematic pattern.
b. Use the DW test (in Excel hand calculation approach) to find out if the error variance is auto correlated.
c. Use the Cochrane-Orcutt two-stage procedure and obtain the estimates of the feasible GLS. If the ϱ (rho) estimated from the Cochrane-Orcutt method differs substantially from that estimated from the DW statistic, which method of estimating ϱ would you choose and why?

The model we will be estimating will be:
close = beta0 + beta1(open) + beta2(high) + beta3(low) + beta4(volume) + beta5(change)

The underlying Economic Theory is that the closing price of a stock depends on intraday price variables (open, high, low) and intraday volume variable. The closing value also depends on the price change in the stock throughout the particular day. This will be proved using econometric tools.

Any data we begin working with, the first step should always be to check if the model is fine. By fine I mean, it should follow all assumptions, and should not have problems like Multicollinearity, Heteroskedasticity, Autocorrelation, among others. Since we have been given questions on Autocorrelation, we will deal with it later.

First, we check for Multicollinearity using Farrar Glauber Test. We reject the null of no multicollinearity using the chi2 test. We view that high, low and open have a strong correlation between each other, which is expected in the price of a share. In the price of a share, all three variables, i.e., opening price of a share, highest share of a price, and lowest price of a share have equal information. Thus, we take only opening price as an independent variable, since variables high and low are not statistically significant, and also multicollinear with open. Just by doing this, the value of chi2 test reduced from 4 digits to a mere 28.69. This is still high, but any other transformation will lead to loss of information.

The model we are estimating now is: close = beta0 + beta1(open) + beta2(volume) + beta3(change)

We now test for heteroskedasticity. We fail to reject the null of homoskedasticity using Breusch Pagan Godfrey Test, thus concluding that we do not have heteroskedasticity. Also, using sktest, we get that the residuals are normally distributed.

Now that we have a model free from heteroskedasticity and multicollinearity, we can proceed with the questions asked.

Estimate a regression model using the given data.

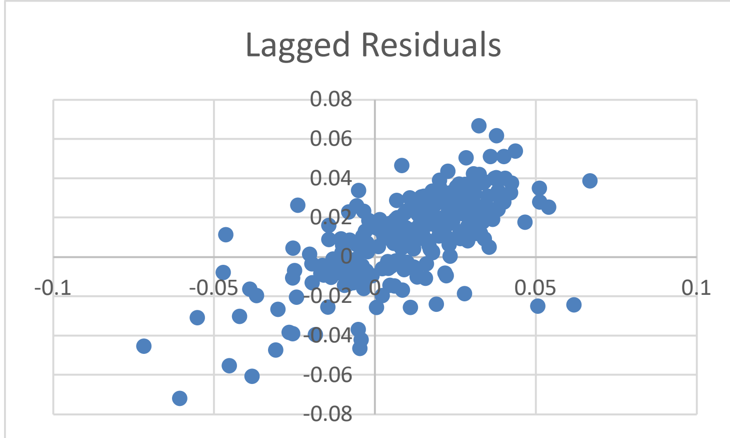
Here we have the model regressed. We get that the coefficient of volume and the constant term are not statistically significant.

```
. reg close open volume change

      Source |       SS           df       MS      Number of obs   =       247
-------------+----------------------------------   F(3, 243)       >  99999.00
       Model |  1852748.42         3  617582.806   Prob > F        =    0.0000
    Residual |  11.4305907       243  .047039468   R-squared       =    1.0000
-------------+----------------------------------   Adj R-squared   =    1.0000
       Total |  1852759.85       246   7531.5441   Root MSE        =    .21689

------------------------------------------------------------------------------
       close | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        open |   .9998789   .0001685  5935.46   0.000     .9995471    1.000211
      volume |   5.83e-06   5.63e-06     1.04   0.301    -5.26e-06    .0000169
      change |   1.000352   .0008773  1140.31   0.000     .9986242     1.00208
       _cons |   .0624436    .232997     0.27   0.789    -.3965079     .521395
------------------------------------------------------------------------------
```

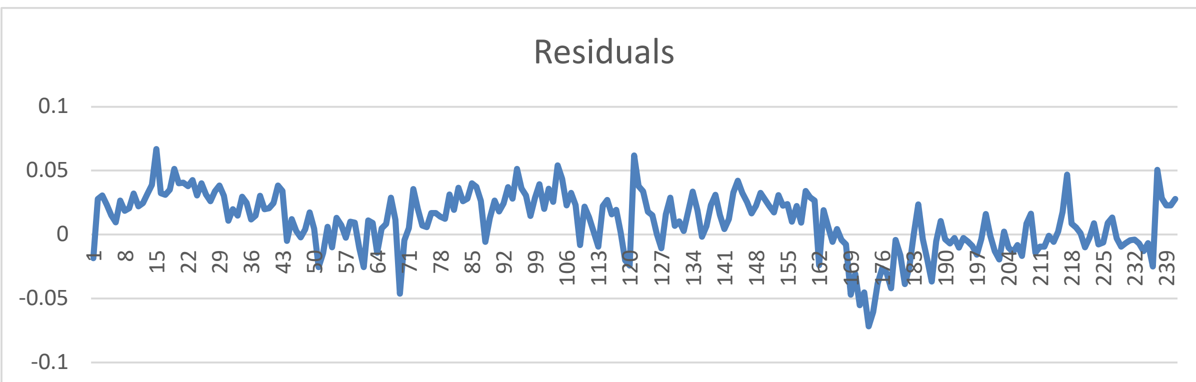
a. Plot the residuals obtained and see if you observe any systematic pattern.



Plotting the residuals on their lag (after removing some outliers), we get that visibly there is positive autocorrelation since the plot is along the line y=x.

Thus, we conclude positive autocorrelation by graphical means.

Also, looking in the graph below, where we have plotted the residuals on a line plot, we can see that clearly the number of runs is too few for a sample with more than 200 observations, so even without doing the Runs Test, we can take an educated guess on the presence of positive autocorrelation by Runs Test. This is further strengthened when we actually run the Runs Test in excel, where we get the number of runs equal to 60, while the lower limit (E(R) – 1.96sd(R)) is 95, confirming that there is indeed significant positive autocorrelation.



b. Use the DW test (in Excel hand calculation approach) to find out if the error variance is auto correlated.

Durbin Watson Test	
ut^2	11.43058
u(ut-u(t-1))^2	23.10993
Test Statistic	2.021764
dl	1.769
du	1.817
4-du	2.183

We get that the value of test statistic = 2.02.
Also, du < test statistic < 4-du.

Thus, by DW test in Excel hand calculation approach, we get that the error variance is not statistically significantly autocorrelated.

This is in an opposite trend to that of the Runs Test. This tells us the fact that we should not rely on only one method for concluding autocorrelation, or for that matter, any issue with the data.

c. Use the Cochrane-Orcutt two-stage procedure and obtain the estimates of the feasible GLS. If the ϱ (rho) estimated from the Cochrane-Orcutt method differs substantially from that estimated from the DW statistic, which method of estimating ϱ would you choose and why?

We have used the Cochrane-Orcutt two-stage procedure to obtain the estimates of the GLS.

We get similar estimates of slope coefficients as when we ran a normal OLS regression on the model.

Also, we got the value of rho equal to 2.02 which is equal to what we got using the Durbin Watson Test.

```
. prais close open volume change, corc

Iteration 0:  rho = 0.0000
Iteration 1:  rho = -0.0110
Iteration 2:  rho = -0.0112
Iteration 3:  rho = -0.0112

Cochrane-Orcutt AR(1) regression with iterated estimates

      Source |       SS           df       MS      Number of obs   =       246
-------------+----------------------------------   F(3, 242)       >  99999.00
       Model |  1869719.23         3  623239.743   Prob > F        =    0.0000
    Residual |  11.4274117       242  .04722071    R-squared       =    1.0000
-------------+----------------------------------   Adj R-squared   =    1.0000
       Total |  1869730.66       245  7631.55369   Root MSE        =     .2173

------------------------------------------------------------------------------
       close | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        open |   .9998841    .000169  5916.34   0.000     .9995511    1.000217
      volume |   6.05e-06   5.72e-06     1.06   0.291    -5.22e-06    .0000173
      change |   1.000407   .0009126  1096.21   0.000     .9986089    1.002204
       _cons |   .0538551   .2353222     0.23   0.819    -.4096863    .5173964
-------------+----------------------------------------------------------------
         rho |  -.0111736
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    = 2.021820
Durbin-Watson statistic (transformed) = 2.000391
```

Notes:

- We do not take the variables high and low since they are the result of random market fluctuations throughout the day. They do not necessarily show us the trends in market, but are more of a show of how the market sentiments were in that day. That is the reason why those variables were dropped.

- Since we got a scenario where Runs Test showed us significant Positive Autocorrelation, and Durbin Watson Test showing us no autocorrelation, we were not able to conclude either of them. In such cases, taking an iterative approach was the way to go. Using the iterative approach, we got the result no autocorrelation.

- Also, since our dataset had 247 observations, we were able to drop the first observation in order to conduct the Cochrane-Orcutt Iterative Procedure. It is the better approach as compared to Prais-Winsten Iterative approach as the latter assumes the missing observation using a transformation, which is not always in the same range as other observations of that variable. Prais-Winston is better in small samples, as it helps us conserve the number of observations.

- Also, the discrepancy between Runs Test and Durbin Watson Test may be because the former looks into broader patterns, while the latter looks into first order autocorrelation. All this proves right now is that there is no statistically significant first order autocorrelation.

The final regression model, after addressing issues of multicollinearity and heteroskedasticity, reliably explains the closing price using opening price, volume, and daily price change, with diagnostic tests indicating no significant autocorrelation in the error variance. All of this while having no heteroskedasticity and reduced multicollinearity makes this a robust model.