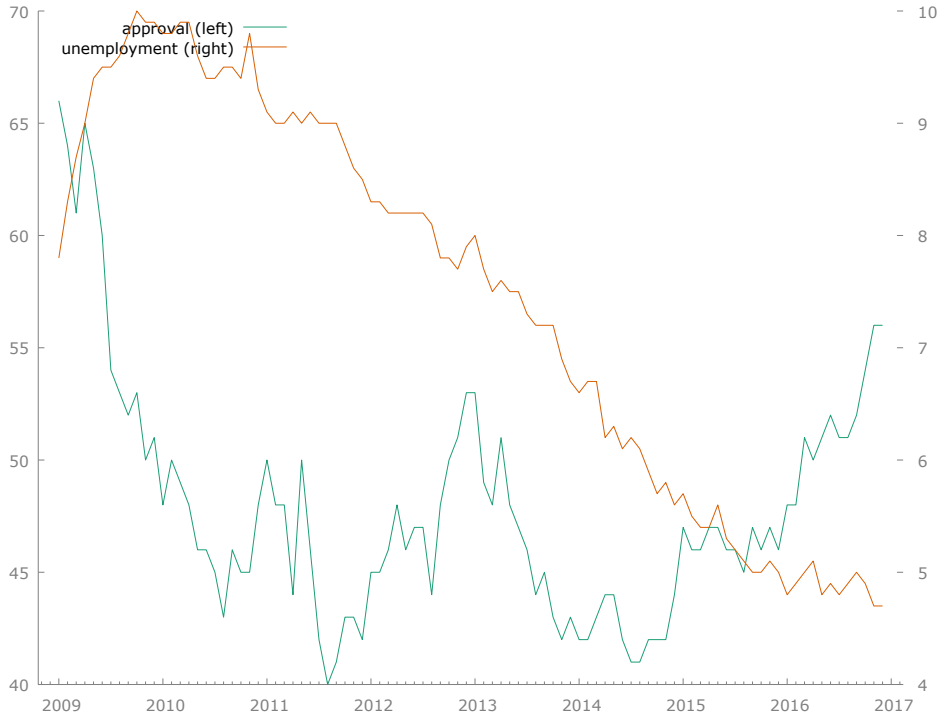
ECON 453
In-Class Exercise 10
November 28, 2023

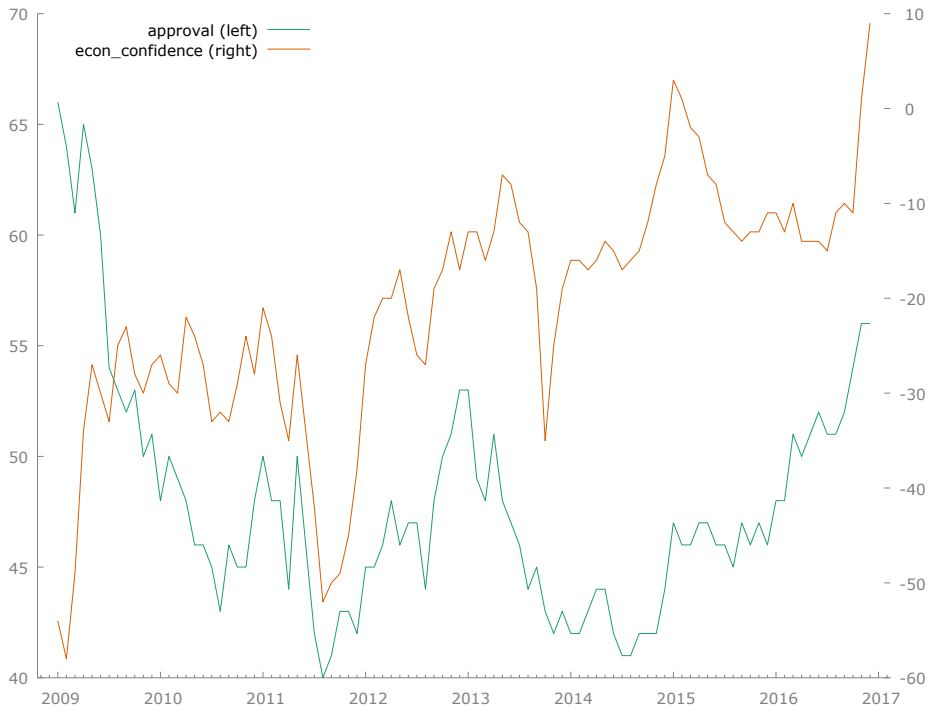
ANSWER KEY

Please download the file "IC10.gdt", a gretl data file. This dataset comes from the Gallup Poll and the Bureau of Labor Statistics. The dataset contains monthly information from January 2009 to December 2016 (96 observations) for the following variables: presidential approval rating (Barack Obama), U.S. unemployment rate, and the Gallup Poll's Economic Confidence Index. The index measures how Americans feel the economy is doing and where it is headed. Negative numbers indicate a pessimistic outlook.

1. First, look at time-plots for the approval and economic variables (**View -> Graph specified vars -> Time series plot**). Look at unemployment and approval, then economic confidence and approval. Comment on trends in the variables, and whether it looks like there is a relationship between approval ratings and the two economic indicators.



The plot above shows the approval rating (green) and the unemployment rate (orange). To me, there doesn't appear to be much of a correlation between the two. Unemployment has a strong downward trend for most of Obama's two terms. This may be due to economic policy choices but is also largely explained by entering office as we were into/entering the "Great Recession". This is one of the trick parts of working with time-series analysis. We would think unemployment would affect the president's approval rating, but there are a lot of other factors that might as well, and there are trends in the data that might make the relationship difficult to identify over a certain time period.



This plot shows the approval rating (green) and the economic confidence index (orange). This one appears to show more of a relationship. The first part is tricky. Obama rode a wave of popularity into office. Like most presidents, this initial welcoming "honeymoon" period tends to wear off quickly. This coincided with the economy tanking and confidence plummeting. After that, however, it mostly looks like the approval rating tends to follow the patterns in the economic confidence index.

2. Run a regression where **approval** is the dependent variable and **unemployment** and **econ_confidence** are the regressors.

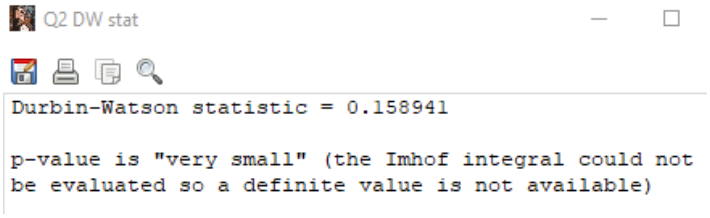
```
Q2: OLS, using observations 2009:01-2016:12 (T = 96)
Dependent variable: approval

                  coefficient   std. error   t-ratio    p-value
  ------------------------------------------------------------------
  const            45.9444       2.60502     17.64     2.03e-031 ***
  unemployment      0.154150     0.437660     0.3522    0.7255
  econ_confidence  -0.0402926    0.0600669   -0.6708    0.5040

Mean dependent var   47.93750   S.D. dependent var   5.377952
Sum squared resid    2696.704   S.E. of regression   5.384869
R-squared            0.018533   Adjusted R-squared  -0.002574
F(2, 93)             0.878044   P-value(F)           0.419011
Log-likelihood      -296.3191   Akaike criterion     598.6382
Schwarz criterion    606.3312   Hannan-Quinn         601.7478
rho                  0.882281   Durbin-Watson        0.158941
```

a. Summarize the findings of this regression

This regression is "hot garbage". It tells us there are no relationships between the approval rating and unemployment or economic confidence and has a negative value for adjusted R^2.

b. Check the Durbin-Watson test and discuss briefly whether there is an issue.

The Durbin-Watson statistic is seen in the bottom of the regression results and is a low value of 0.159. Remember that a "good" value of this statistic would be between about 1.5 and 2.5 (2 is ideal). Low values (close to 0) or high values (close to 4) indicate a problem of serial correlation. We can click the Tests -> Durbin-Watson p-value to confirm our worst fears.



It is official, our nightmare has come true. The null hypothesis in this test is that there is no problem with serial correlation. A p-value that is "very small" means we need to reject the null.

3. Let's work on accounting for trends in our regression. Add a time trend variable to your dataset (**Add -> Time-trend**)
    a. Regress the **approval** variable on the time trend variable (only the time trend variable). Save the fitted values from this regression (**Save -> Fitted values** from the regression results window). Call these **linear_fit**.

```
Q3 linear:
OLS, using observations 2009:01-2016:12 (T = 96)
Dependent variable: approval

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       50.5809       1.06703     47.40     1.98e-067 ***
  time        -0.0545035    0.0191023   -2.853    0.0053     ***

Mean dependent var   47.93750   S.D. dependent var   5.377952
Sum squared resid    2528.630   S.E. of regression   5.186552
R-squared            0.079703   Adjusted R-squared   0.069913
F(1, 94)             8.140988   P-value(F)           0.005323
Log-likelihood      -293.2302   Akaike criterion     590.4603
Schwarz criterion    595.5890   Hannan-Quinn         592.5334
rho                  0.892105   Durbin-Watson        0.156287
```

The linear trend model tells us that there is a significant negative trend in the approval rating during Obama's time in office.

    b. Regress the **approval** variable on the time trend variable and the squared version of the time trend variable (a quadratic trend model).
        i. Save the fitted values from this regression. Call these **quadratic_fit**.

```
Q3 quad: OLS, using observations 2009:01-2016:12 (T = 96)
Dependent variable: approval

             coefficient   std. error    t-ratio    p-value
  ----------------------------------------------------------
  const       59.3104       1.10151      53.84    6.67e-072 ***
  time        -0.588962     0.0524168   -11.24    5.17e-019 ***
  sq_time      0.00550988   0.000523563   10.52    1.61e-017 ***

Mean dependent var   47.93750   S.D. dependent var   5.377952
Sum squared resid    1154.167   S.E. of regression   3.522840
R-squared            0.579940   Adjusted R-squared   0.570907
F(2, 93)             64.19849   P-value(F)           3.05e-18
Log-likelihood      -255.5838   Akaike criterion     517.1677
Schwarz criterion    524.8607   Hannan-Quinn         520.2773
rho                  0.815740   Durbin-Watson        0.325975
```

By comparing the linear and the quadratic trend regression results, we can see that the quadratic model is a much better fit. Both the linear and quadratic versions of the "time" variable are highly significant, and our adjusted R^2 jumped from 0.07 to 0.57. The coefficients tell us the trend started negative but flattened out and turned positive as more time passed.

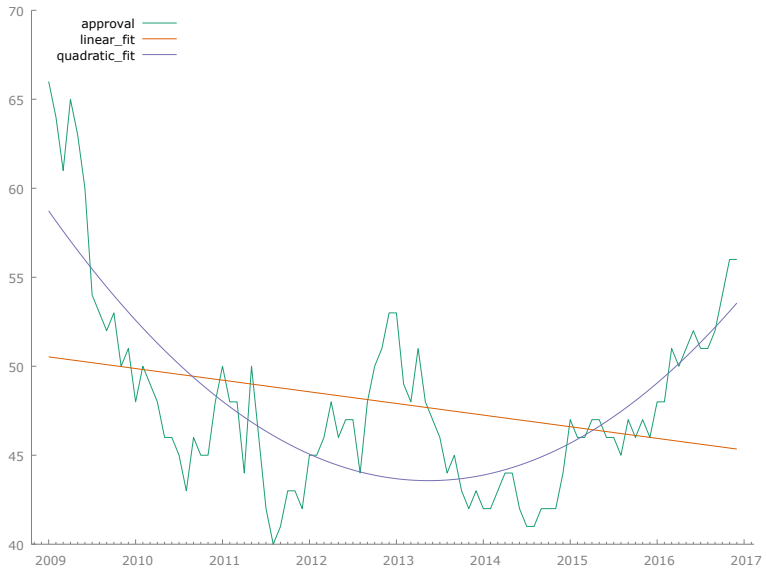
ii. What does your quadratic model predict President Obama's approval rating would be in January 2017?

The equation can be written as: $\hat{y}_t = 59.31 - 0.59 * t + 0.0055 * t^2$. The time variable increases by 1 each month. It starts as 1 in January 2009 and ends at 96 in December 2016. This means January 2017 would be t=97 in the sequence. We can predict approval for this month by plugging t=97 into our equation.

$$\hat{y}_t = 59.31 - (0.59 * 97) + (0.0055 * (97^2) = 53.83$$

I predict Obama's approval rating in January 2017 would be 53.83%.

c. Compare the results of the two models. View a time-plot with approval ratings and the fitted values from each of the models on one plot. Which is the better way to model the trend in approval ratings?



I discussed this previously, but the quadratic trend model is a much better fit based on the p-values and adjusted R^2. The plot confirms how much better the quadratic model is in this case.

4. Next, run a regression using **approval** as the dependent variable and the following regressors: time trend, squared time trend, **unemployment**, and **econ_confidence**.

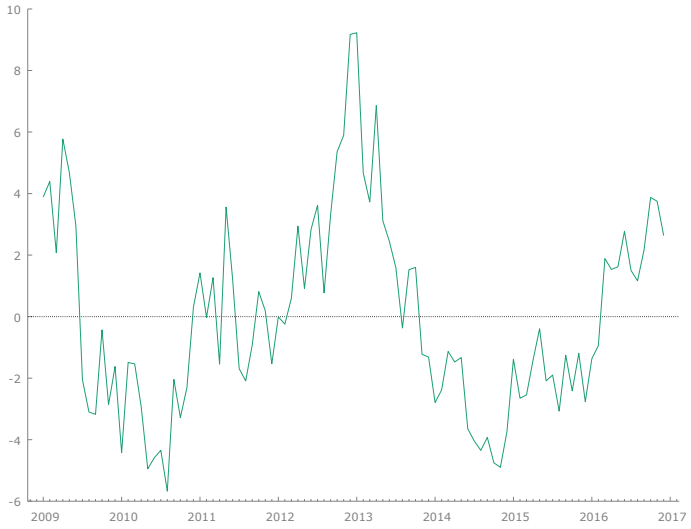
```
Q4: OLS, using observations 2009:01-2016:12 (T = 96)
Dependent variable: approval

                    coefficient   std. error    t-ratio    p-value
  ------------------------------------------------------------------
  const             90.1939       8.12010       11.11    1.28e-018 ***
  time              -0.699434     0.0548240    -12.76    5.62e-022 ***
  sq_time            0.00454580   0.000570560    7.967   4.48e-012 ***
  unemployment      -2.69000      0.790575      -3.403   0.0010    ***
  econ_confidence    0.118600     0.0398059      2.979   0.0037    ***

Mean dependent var   47.93750   S.D. dependent var   5.377952
Sum squared resid    961.0165   S.E. of regression   3.249711
R-squared            0.650237   Adjusted R-squared   0.634863
F(4, 91)             42.29412   P-value(F)           5.34e-20
Log-likelihood      -246.7930   Akaike criterion     503.5860
Schwarz criterion    516.4077   Hannan-Quinn         508.7687
rho                  0.820686   Durbin-Watson        0.347532
```

a. Save the residuals from this regression. ***Fine.***

b. Comment on the findings of the model. What does this model predict about the role of economic factors in the president's approval ratings?

The model tells us that, as unemployment increases, the president's approval rating drops. There is a positive and significant relationship between economic confidence and the approval ratings. These are the effects I would have expected.

c. View a time plot of the residuals from your regression.



i. Does this seem to indicate a serial correlation problem?

Oh No! These are indicative of a problem. We do not want to see what look like "trends" in our regression residuals. Periods of time where all the residuals are positive (or all negative) are an example of indicators of serial correlation.

ii. Do you notice anything regarding the dates that might help us explain/model trends in presidential approval ratings?

One thing we can do while visualizing the residuals against time is look for events/patterns we might be missing in our model. For example, the residuals are all high at the beginning of Obama's presidency. This means the approval ratings were higher than expected. This means we might want to account for the "honeymoon" period in some way. There is a spike in May 2011. This is when Osama Bin Laden was killed. There is a large spike late in 2012/early 2013. This is when Obama was campaigning and was reelected.

d. Check the correlogram of the residuals from your model and the Durbin-Watson statistic from your regression. Are we gonna have a problem here?

If you saved the residuals (as I demanded earlier in the problem), then you can go to your main gretl window. Find where the residuals are, right-click on them, and choose correlogram. You will get the results below. Ideally, we would see no significant correlations here. We see correlations, which means we are, indeed, gonna have a problem here. The quadratic trend model has helped us identify more realistic relationships between economic variables and presidential approval, but it has not dealt with our serial correlation problem.

```
Autocorrelation function for uhat5
***, **, * indicate significance at the 1%, 5%, 10% levels
using standard error 1/T^0.5

 LAG      ACF          PACF         Q-stat. [p-value]

   1   0.8148  ***   0.8148 ***     65.7411  [0.000]
   2   0.7035  ***   0.1179        115.2710  [0.000]
   3   0.6173  ***   0.0471        153.8145  [0.000]
   4   0.5362  ***  -0.0046        183.2134  [0.000]
   5   0.4513  ***  -0.0452        204.2744  [0.000]
   6   0.3657  ***  -0.0567        218.2553  [0.000]
   7   0.2226  **   -0.2382 **     223.4935  [0.000]
   8   0.1916  *     0.1780 *      227.4165  [0.000]
   9   0.1447       -0.0144        229.6816  [0.000]
  10   0.0731       -0.0902        230.2654  [0.000]
  11   0.0027       -0.0662        230.2663  [0.000]
  12  -0.0720       -0.1002        230.8466  [0.000]
```

5. Let's try a first-differences model instead. Create the differenced variables for approval, unemployment, and economic confidence (**Add -> First differences of selected variables**).
    a. Regress the differences in approval on the differences in unemployment and economic confidence.

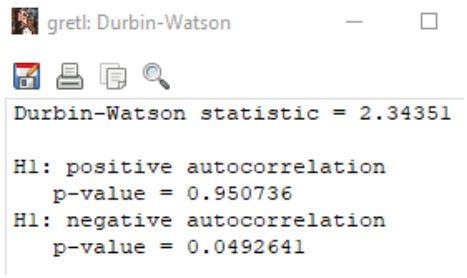
```
Q5: OLS, using observations 2009:02-2016:12 (T = 95)
Dependent variable: d_approval

                     coefficient   std. error   t-ratio   p-value
  ------------------------------------------------------------------
  const                -0.247383    0.201959     -1.225    0.2237
  d_unemployment       -1.38330     1.12118      -1.234    0.2204
  d_econ_confidence     0.146240    0.0392878     3.722    0.0003  ***

Mean dependent var  -0.105263   S.D. dependent var   2.049773
Sum squared resid    338.6784   S.E. of regression   1.918668
R-squared            0.142472   Adjusted R-squared   0.123830
F(2, 92)             7.642574   P-value(F)           0.000850
Log-likelihood      -195.1799   Akaike criterion     396.3598
Schwarz criterion    404.0215   Hannan-Quinn         399.4557
rho                 -0.173784   Durbin-Watson        2.343513
```

    b. Comment on the findings of this model.

This model tells us that unemployment has a negative, but not significant relationship with Obama's approval rating. Economic confidence has a significant positive effect on approval. The second one makes sense, and the first one doesn't surprise based on what we saw at the beginning of the assignment (unemployment mostly decreased throughout Obama's two terms as president).

    c. Based on the Durbin-Watson test, do we have a problem with serial correlation in this model?

The Durbin-Watson statistic (2.34) is in the sweet spot we would like to see (between about 1.5 and 2.5). We can check by examining the p-value. This actually tells us that we are on the line of having a problem with negative autocorrelation in this model. We can check with the other types of autocorrelation tests in cases like this where we are on the line.

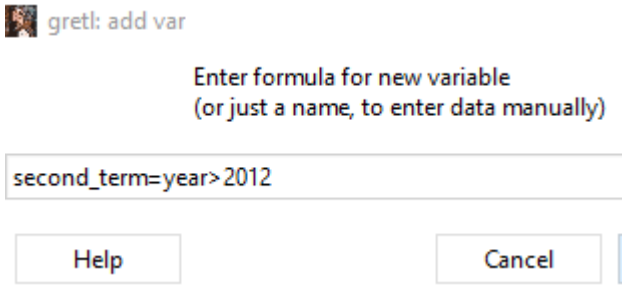


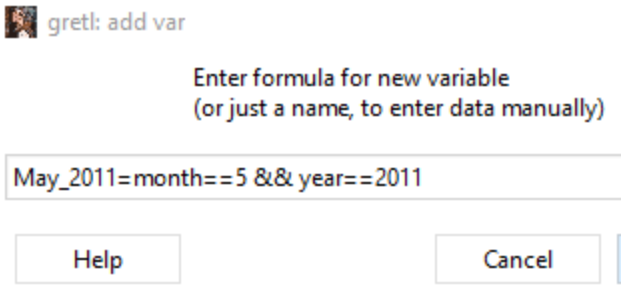
d. Check other tests of autocorrelation (from the regression results select **Tests -> Autocorrelation**). You can use the default option of 12 lags. The null in each of these tests is "no problem". Overall, do we seem to have a problem with serial correlation?

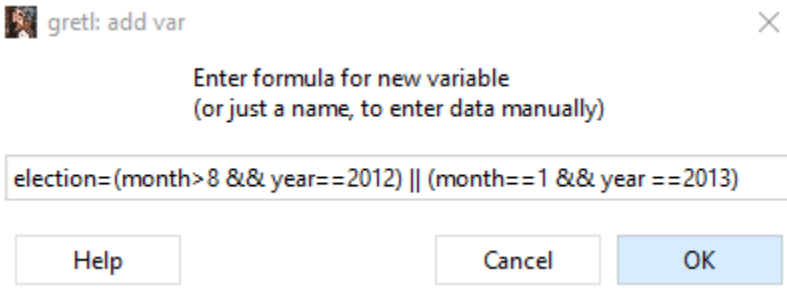
```
Breusch-Godfrey test for autocorrelation up to order 12
OLS, using observations 2009:02-2016:12 (T = 95)
Dependent variable: uhat

                     coefficient   std. error   t-ratio    p-value
  ------------------------------------------------------------------
  const               -0.0203993    0.200992    -0.1015    0.9194
  d_unemployment      -0.461454     1.18561     -0.3892    0.6982
  d_econ_confidence    0.0242305    0.0427724    0.5665    0.5726
  uhat_1              -0.178625     0.112092    -1.594     0.1150
  uhat_2              -0.0119137    0.114413    -0.1041    0.9173
  uhat_3               0.0863934    0.117790     0.7335    0.4654
  uhat_4               0.0879780    0.117633     0.7479    0.4567
  uhat_5               0.130210     0.115665     1.126     0.2636
  uhat_6               0.238424     0.116534     2.046     0.0440   **
  uhat_7              -0.196711     0.114329    -1.721     0.0892   *
  uhat_8              -0.0276831    0.117431    -0.2357    0.8142
  uhat_9               0.0652488    0.118371     0.5512    0.5830
  uhat_10              0.0914628    0.118748     0.7702    0.4434
  uhat_11              0.00197234   0.121087     0.01629   0.9870
  uhat_12             -0.0497265    0.120655    -0.4121    0.6813

  Unadjusted R-squared = 0.161916

Test statistic: LMF = 1.287988,
with p-value = P(F(12,80) > 1.28799) = 0.242

Alternative statistic: TR^2 = 15.382045,
with p-value = P(Chi-square(12) > 15.382) = 0.221

Ljung-Box Q' = 17.4912,
with p-value = P(Chi-square(12) > 17.4912) = 0.132
```

Based on the p-values of these tests (at the bottom of the results), we should not be too concerned about a serial correlation problem. Great news, everybody!

6. Let's try to account for some unique time periods. Create dummy variables for the second term (starting January 2013), the month of May in 2011, and the re-election period (let's call it September 2012 to January 2013).







a. Run a regression using first-differences of approval as the dependent variable, and the following explanatory variables: first-differences of unemployment and economic confidence, your 3 time-period dummies.

```
Q6: OLS, using observations 2009:02-2016:12 (T = 95)
Dependent variable: d_approval

                     coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------------
  const               -0.654579    0.277739    -2.357    0.0206  **
  second_term          0.545054    0.386260     1.411    0.1617
  May_2011             5.48160     1.86730      2.936    0.0042  ***
  d_unemployment      -0.988278    1.08302     -0.9125   0.3640
  d_econ_confidence    0.119351    0.0379170    3.148    0.0022  ***
  election             1.99162     0.847675     2.350    0.0210  **

Mean dependent var  -0.105263   S.D. dependent var   2.049773
Sum squared resid    292.7485   S.E. of regression   1.813645
R-squared            0.258766   Adjusted R-squared   0.217123
F(5, 89)             6.214002   P-value(F)           0.000055
Log-likelihood      -188.2574   Akaike criterion     388.5148
Schwarz criterion    403.8381   Hannan-Quinn         394.7066
rho                 -0.111646   Durbin-Watson        2.219830
```

b. Comment on what we learn from these results.

We learn there is no overall difference between the first and second terms in terms of approval ratings. We see a significant spike in approval ratings when the killing of Osama Bin Laden occurred, as well as a spike during the reelection period. As before, our unemployment coefficient is negative but not at all significant, and economic confidence has a positive and significant relationship with approval.