TIME SERIES PROJECT

Instructor: Narek OHANYAN
Student: Elina SIMONYAN

FEBRUARY 2023

Introduction

Companies around the world spend well over $600 billion per year on ads—and digital advertising specifically is set to reach $876 billion by 2024. But why is advertising important? Or is it really important? Tesla spends $0 on advertising, according to a new report from BrandTotal. But competitors such as Toyota, BMW, Porsche, and Ford spend heavily on the major social platforms: Facebook, YouTube, Instagram, and Twitter.
This small project examines the effect of marketing expenses on the sales. To study whether the companies need to spend much on their ads quarterly sales of ADIDAS are selected, from Q1 2000- Q1 2017. The data is collected from Kaggle` an online community platform, as well as various internet sources to find advertisement expenses and to check sales/revenues with income statements of the company.

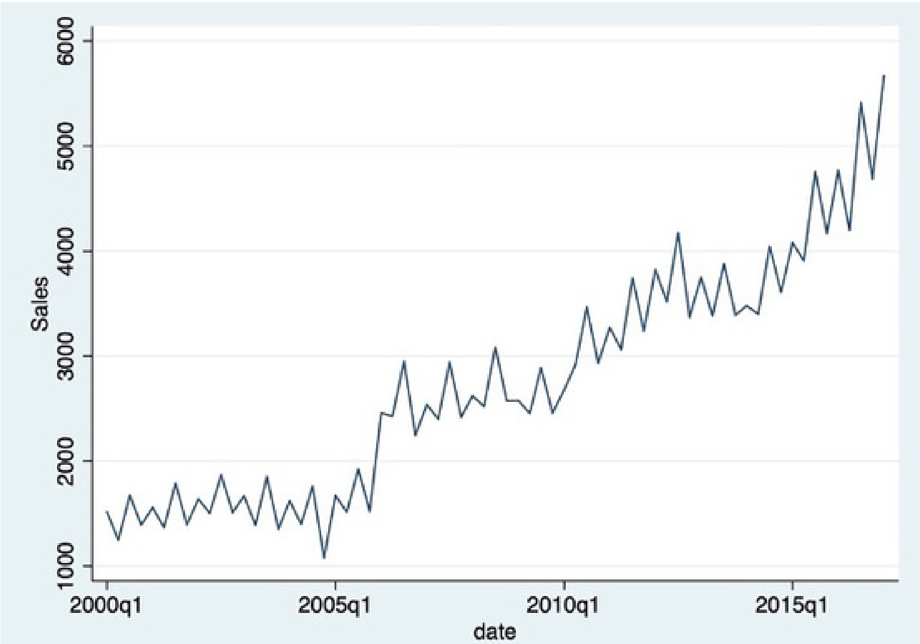
In this project we initially have 2 variables:

Sales -Which are quarterly sales of the company Q1 2000- Q1 2017.
Adv_exp – which shows quarterly advertisement/ marketing expenses Q1 2000- Q1 2017.

Let's check whether they are stationary or not, and if not, what are steps we need to do.

Determining the presence of stationarity

"Stationarity" implies that data spanned over different time periods should have a constant mean and variance.
Creating a visual plot of data is the first step in time series analysis. Graphical representation of data helps understand it better.
When plotting the graph of the sales (graph 1), it is obvious that series are not stationary.



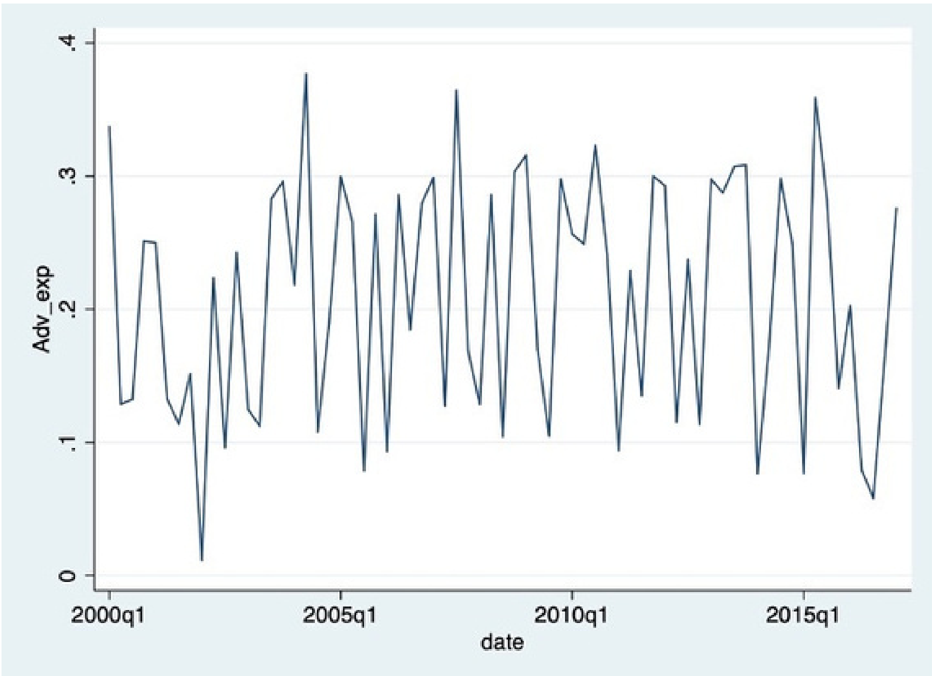
expressed in mln. USD

Obviously from the graph Sales time series has been increasing with minor fluctuations. Thus, time series cannot have a constant mean and variance. Therefore, the primary assumption of time series, i.e., stationarity, is missing.

The variable "Sales" here has an upward trend with non-constant mean and variance. Thus, Sales time series are non-stationary.

But at the same time marketing expenses are more or less stationary and obviously have seasonality.

However, the graph is only an initial step in the formal stability testing process.



Augmented Dickey Fuller test

The Dickey Fuller test helps to examine the stationarity of time series data. An important assumption of this test is that the error term is uncorrelated. Therefore, the Augmented Dickey Fuller test is performed first. It checks for correlation in the error term by adding lags.

TO CHECK THE STATIONARITY OF THE SERIES WE NEED TO CHECK THE FOLLOWING HYPOTHESIS:

H0: Series are non-stationary

H1: Series are stationary:

Let's perform the test on stata.

```
. dfuller Sales

Dickey-Fuller test for unit root                   Number of obs   =        68

                              ——————— Interpolated Dickey-Fuller ———————
                  Test        1% Critical       5% Critical      10% Critical
               Statistic          Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.007            -3.555            -2.916            -2.593
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.7507
```

To check for stationarity, we need to focus on only two values of the result: Z(t) and the Mackinnon p-value for Z(t). For time series data to be stationary, Z(t) must be a large negative number. The p-value must be significant at at least the 5% level. In this test, none of these conditions are met. Therefore, the null hypothesis, that is, the time series data are non-stationary, cannot be rejected. And since the Sales time series is non-stationary, no further analysis can be performed on it.

We can run the same test with another command in stata as well and get the results, which also fails to reject the null.

```
. dfuller Sales, noconstant lag(0)

Dickey-Fuller test for unit root                   Number of obs   =        68

                               ---------- Interpolated Dickey-Fuller ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             0.605            -2.613            -1.950            -1.610
```

While testing stationarity of the Advertisement expenses we can see that the series are stationary, as we get a large negative Z(t), and .the p-value is significant at at the 5% level.

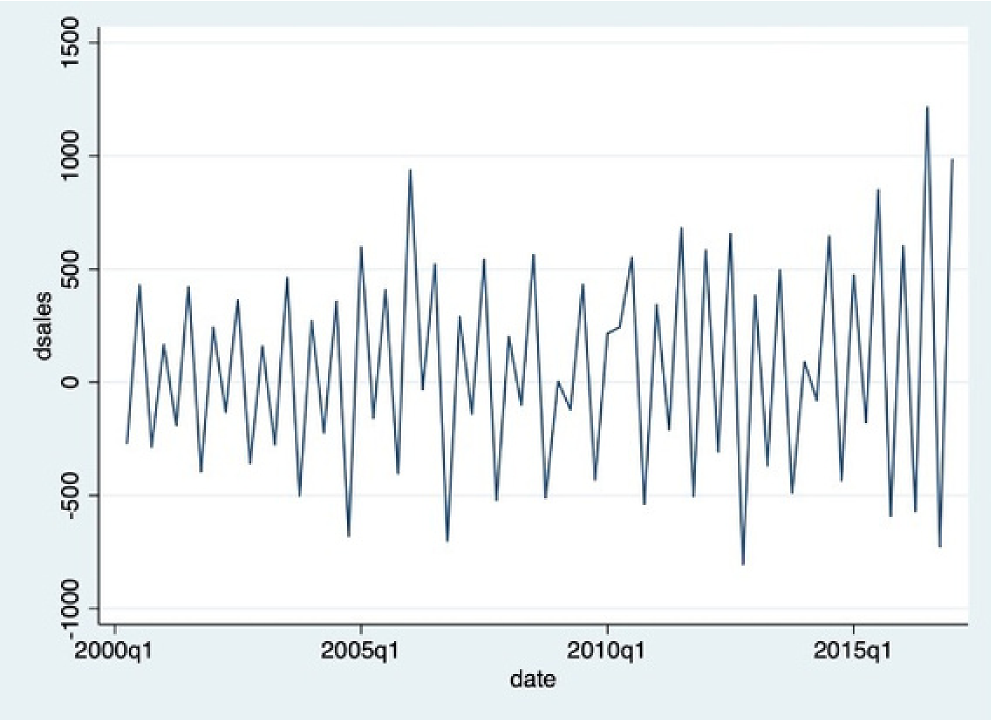
```
. dfuller Adv_exp

Dickey-Fuller test for unit root                   Number of obs   =        68

                               ---------- Interpolated Dickey-Fuller ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -8.950            -3.555            -2.916            -2.593
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

In the above Dickey Fuller tests, lags were excluded, assuming the error term is uncorrelated. If lags are included, stationarity of Sales time series can change. To do the same, we need to perform Augumented Dickey Fuller test. The results after performing were the follows: again, Z(t) value does not have any large negative number. Also p-value is insignificant. Thus, again the null hypothesis of Dickey Fuller test, which states that the time series data is non-stationary, cannot be rejected. Therefore, time series Sales is non-stationary even after taking lags for correlated error terms.

First differencing series

If a time series has a unit root problem, the first difference of such time series is ‘stationary’. Therefore, the solution here is to take the first difference of the Sales time series. The first difference of a time series is the series of changes from one period to the next. If Yt denotes the value of the time series Y at period t, then the first difference of Y at period t is equal to Yt-Yt-1.

Graphicly taking the change in Sales we have this result:



From the graph it is obvious that Sales time series has been diminished to around zero. This implies that this time series can have a constant mean and variance.

However, a mere graphical representation is an intuitive step thus, performing the formal tests of stationarity.

To examine the presence of stationarity, we need to review two values; ‘Z(t)’ and Mackinnon p-value for ‘Z(t)’.

```
. dfuller dsales

Dickey-Fuller test for unit root                   Number of obs   =        67

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -24.463            -3.556            -2.916            -2.593
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

Here from the graph it can be seen that the ‘Z(t)’ value is -24.463 which is a large negative number (as compared to ‘z(t)’ for ‘gdp’) and the p-value is also found significant. Thus, the null hypothesis of the Dickey-Fuller test is rejected. Therefore, the first differenced time series Sales is stationary.

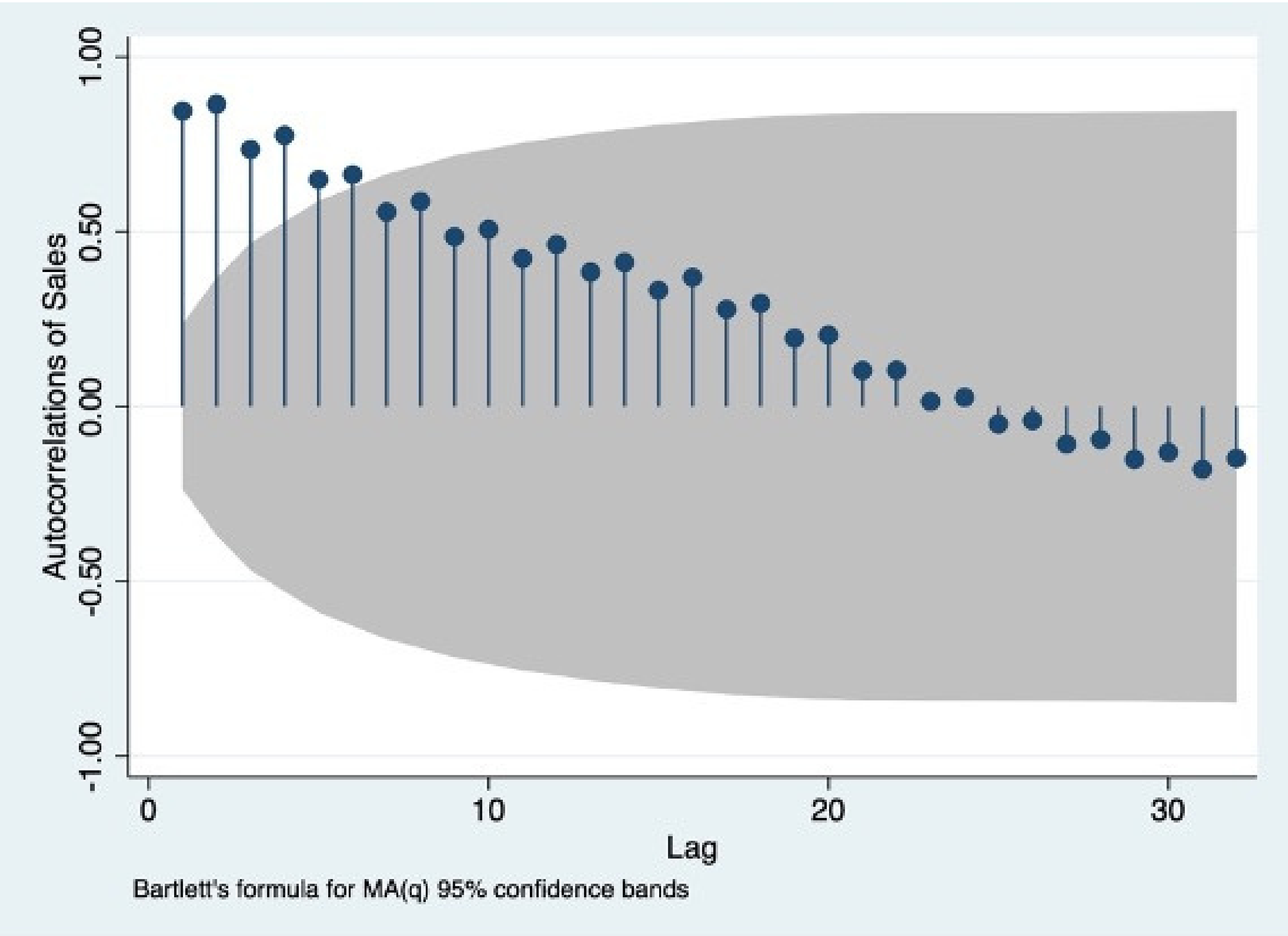
As Sales can be made stationary by taking the first difference, we can conclude that the variable is integrated of order one.

Test for autocorrelation

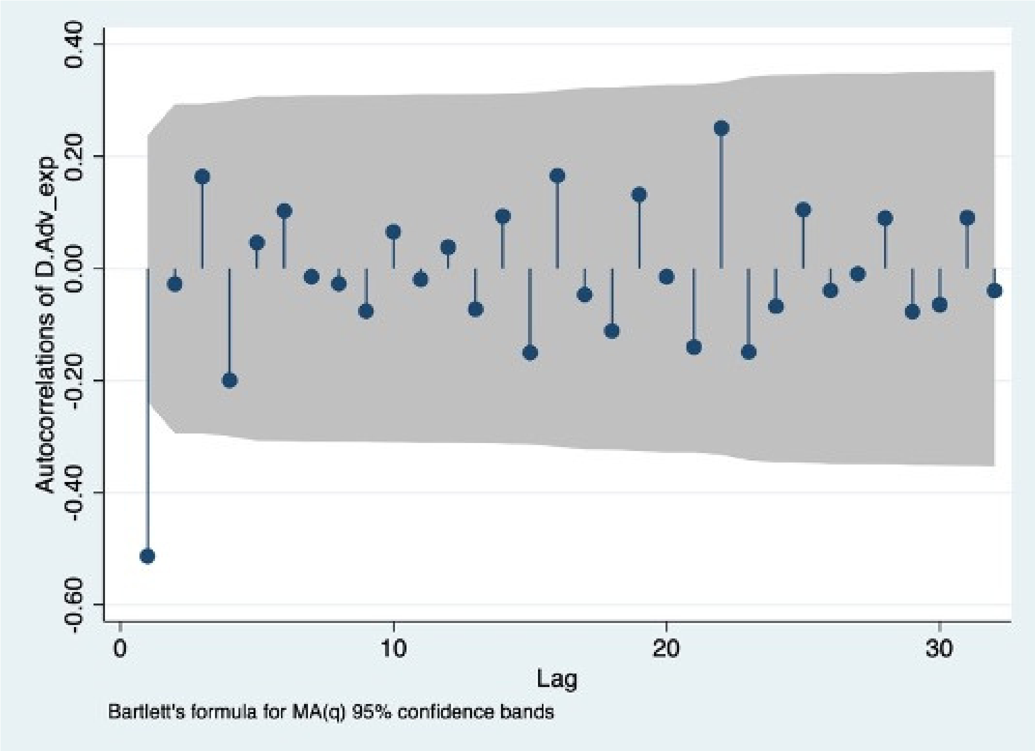
It is one of the main assumptions of the OLS estimator according to the Gauss-Markov theorem that in a regression model:

Cov(ϵ_(i,) ϵ_j)=0 ∀i,j,i≠j,
where Cov is the covariance and ϵ is the residual.

The presence of autocorrelation in the data causes and correlates with each other and violates the assumption, showing bias in the OLS estimator. It is therefore important to test for autocorrelation and apply corrective measures if it is present.
Let’s look at the correlogram:



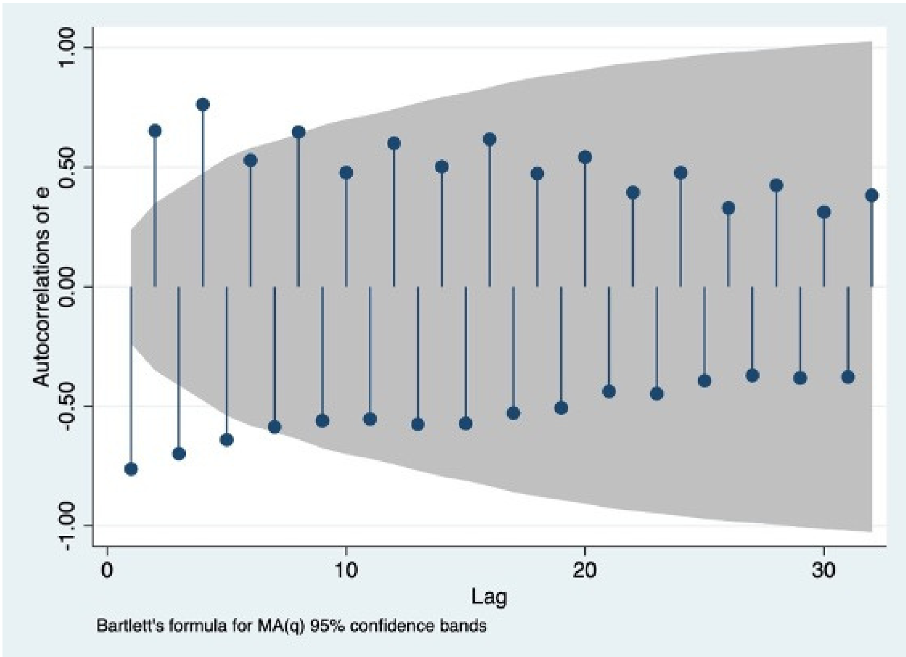
We have statistically significant autocorrelation coefficients up to lag 6.



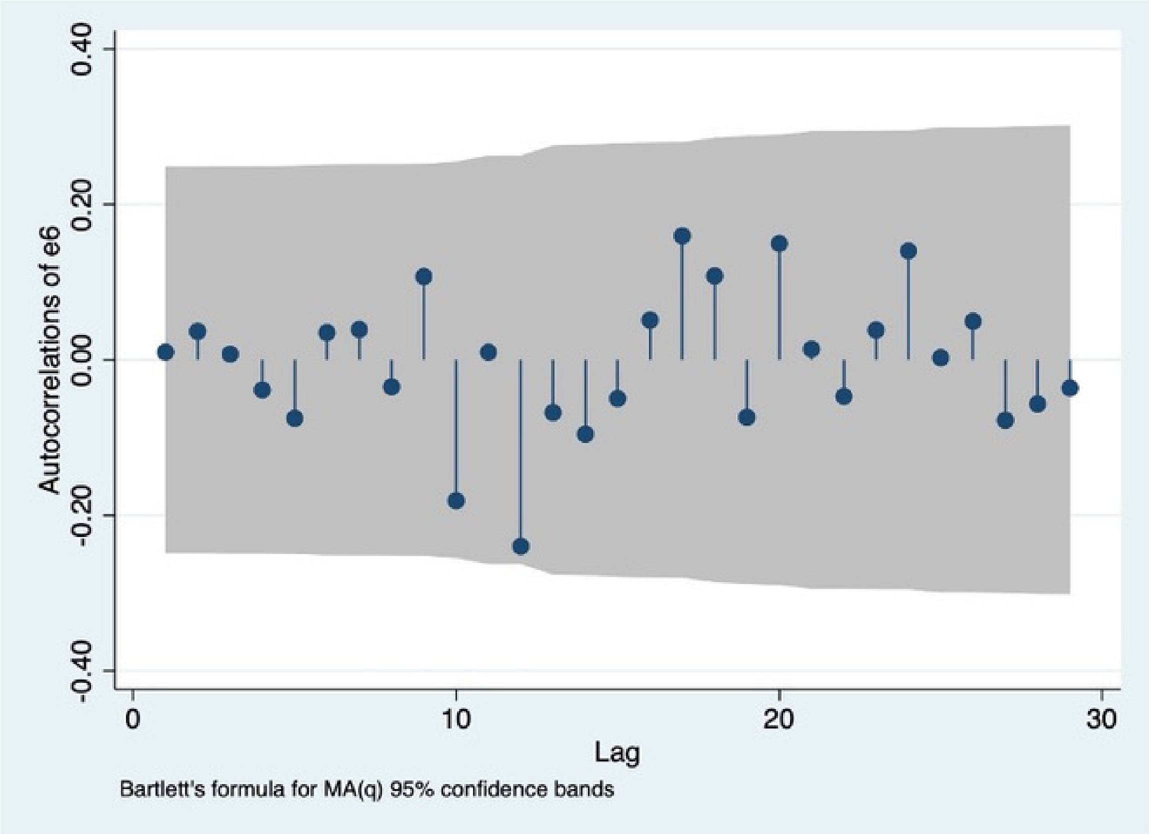
While looking at the changes of marketing expenses we have statistically significant autocorrelation coefficients up to lag 1.

Let’s inspect the residuals and check weather they are autocorrelated over time or not.

We can conclude that standard errors are incorrect. So, we need to use HAC adjusted errors.



To eliminate autocorrelation let’s use ARDL to include 6 lags of dependent variable and 1 lag of independent variable. Then me need to predict residuals and plot the correlogram to check whether we managed to co eliminate autocorrelation or not. Let’s look at the new correlogram:



So we can conclude that we have managed to eliminate autocorrelation in the error terms.

Let's use LM test to check whether there is autocorrelation or not.

The hypothesis in this case is:

- Null hypothesis: There is no serial correlation.
- Alternative Hypothesis: There is a serial correlation.

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
1	**0.063**	**1**	**0.8023**

H0: no serial correlation

Since from the above table, chi2 is more than 0.05 or 5%, the null hypothesis cannot be rejected. In other words, there is not a serial correlation between the residuals in the model.

Selecting lags with AIC and BIC

Looking at different results for AIC and BIC, we can conclude, that we need 6 lags for dSales (both AIC and BIC) and 1 lag for Adv_exp. The 1st Table shows the final option. Below you can see other results.

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	**62**	**-472.2694**	**-408.7127**	**9**	**835.4255**	**854.5697**

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	**65**	**-494.0218**	**-446.207**	**6**	**904.414**	**917.4604**

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	**67**	**-508.7628**	**-472.6912**	**4**	**953.3825**	**962.2012**

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

Final result

```
. reg dsales L.dsales L2.dsales L3.dsales L4.dsales L5.dsales L6.dsales Adv_exp L.Adv_exp

      Source |       SS           df       MS      Number of obs   =        62
-------------+----------------------------------   F(8, 53)        =     44.85
       Model |  13071930.9         8  1633991.36   Prob > F        =    0.0000
    Residual |  1930982.05        53  36433.6235   R-squared       =    0.8713
-------------+----------------------------------   Adj R-squared   =    0.8519
       Total |  15002912.9        61  245949.392   Root MSE        =    190.88

------------------------------------------------------------------------------
      dsales |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dsales |
         L1. |  -.1848967   .1323352    -1.40   0.168    -.4503276    .0805342
         L2. |   .1780411   .1334204     1.33   0.188    -.0895664    .4456485
         L3. |  -.3323237   .1119069    -2.97   0.004    -.5567805    -.107867
         L4. |   .6065506   .1125728     5.39   0.000     .3807581    .8323431
         L5. |  -.0959548   .1396135    -0.69   0.495    -.3759841    .1840744
         L6. |  -.3420159   .1395925    -2.45   0.018     -.622003   -.0620287
             |
     Adv_exp |
         --. |   265.1527   215.0825     1.23   0.223    -166.2482    696.5536
         L1. |   171.8964   218.0231     0.79   0.434    -265.4025    609.1952
             |
       _cons |   67.83156   30.55658     2.22   0.031     6.542817    129.1203
------------------------------------------------------------------------------
```

So, the final model gives this regression results. As already mentioned, the aim is to estimate how the advertisement expenses effects the sales.

As we can see from the result, change in sales depends on the advertisement expenses, both from the same period as well as 1 period apart. Obviously, advertisement expenses positively effects the sales. But we get both high R-square and high p-value. It means that the model explains a lot of variation within the data but is not significant.