STATISTICAL TESTS FOR AUTOCORRELATION AND RESOLVING AUTOCORRELATION PROBLEM

MOHIT SAINI (21HS10028)

1) PROBLEM STATEMENT

India's Consumer Spending and GDP have been chosen for the regression model. The aim is to identify whether Consumer Spending is related to GDP by using linear regression and examine the model for the problem of autocorrelation and apply remedial measures if necessary.

2) MODEL SPECIFICATION

The variable “spending” is the yearly consumer spending of India and the variable “gdp” is the yearly India’s Gross Domestic Product. Yearly data from 1960-2021 is taken.

3) SOURCE OF DATA

The data has been taken from:
https://www.macrotrends.net/

4) MEASUREMENT OF VARIABLES

Both consumer spending and GDP are in billion US$.

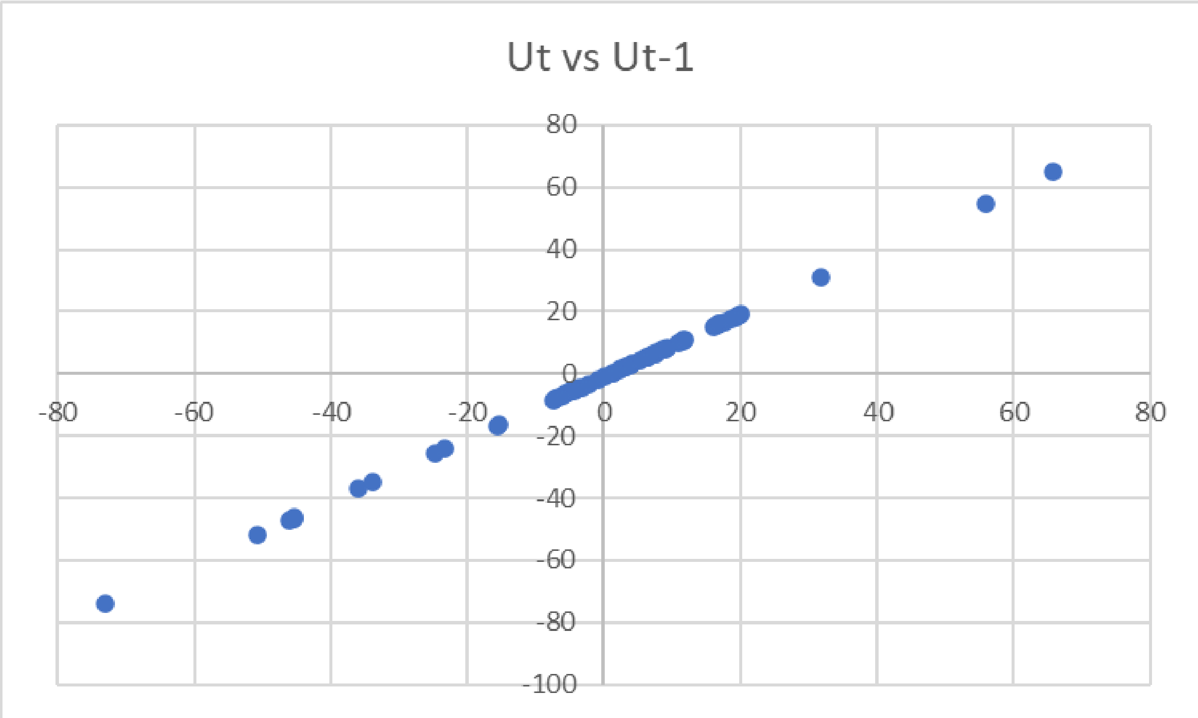
5) ESTIMATION PROCESS FOLLOWED

i) **Regression of spending on gdp**

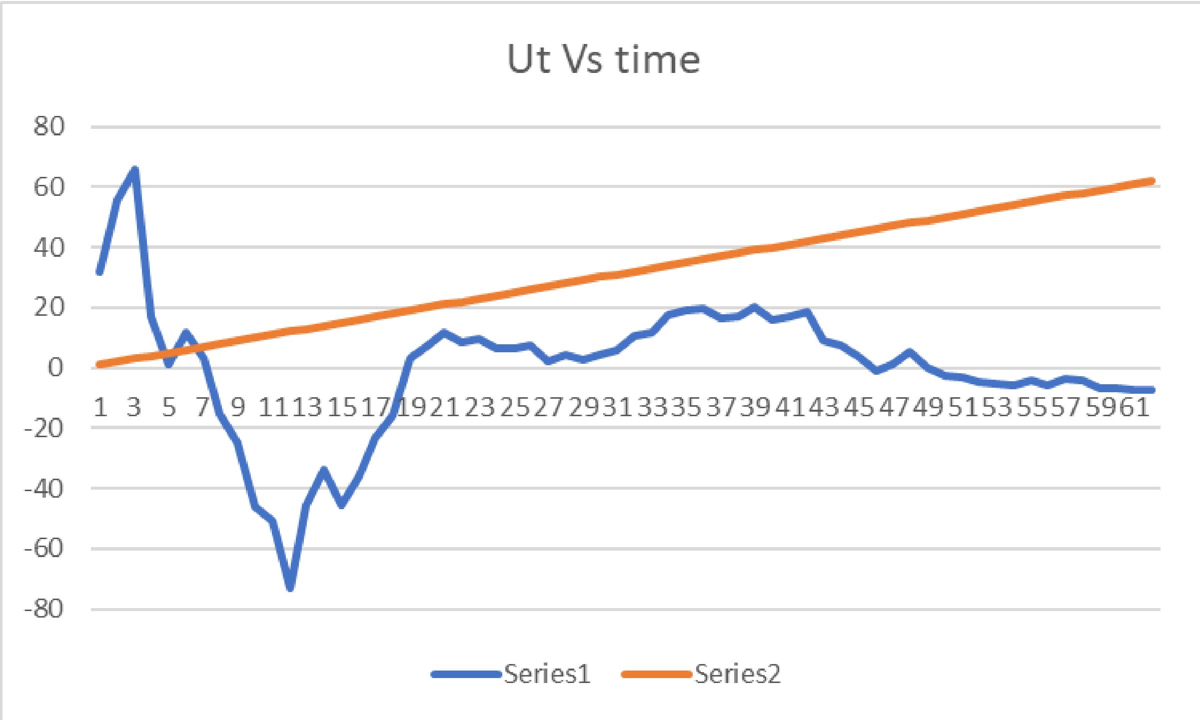
```
. reg spending gdp

      Source |       SS           df       MS      Number of obs   =        62
-------------+----------------------------------   F(1, 60)        =  29421.23
       Model |  15431749.4         1  15431749.4   Prob > F        =    0.0000
    Residual |  31470.6448        60  524.510746   R-squared       =    0.9980
-------------+----------------------------------   Adj R-squared   =    0.9979
       Total |  15463220.1        61  253495.411   Root MSE        =    22.902

------------------------------------------------------------------------------
    spending |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |   .5798843   .0033807   171.53   0.000     .5731219    .5866468
       _cons |   18.25838   3.747743     4.87   0.000     10.76178    25.75499
------------------------------------------------------------------------------
```

ii) **Checking for problems for Autocorrelation-**

A) Graphical Tests-



The graph between Ut vs Ut-1 shows positive upward sloping. It means that there can be a problem with positive autocorrelation.



The line graph between ut and time shows changes in the sign of the residual (ut), very few times, we suspect positive autocorrelation.

B) Runs Test

runtest u, mean

N(u <= 5.19137228689e-08) = 25

N(u > 5.19137228689e-08) = 37

obs = 62

N(runs) = 6

z = -6.61

Prob>|z| = 0

Runs test shows the problem of positive autocorrelation.

C) Durbin Watson Test

estat dwatson

Durbin-Watson d-statistic(2, 62) = 0.2220043

DW test shows positive autocorrelation.

d-statistic (0.2220043) value is in between 0 and dL. So, positive autocorrelation.

D) Breusch–Godfrey Test

```
40 . estat bgodfrey
```

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	**47.276**	**1**	**0.0000**

H0: no serial correlation

BG test rejects the null hypothesis, which means the problem of autocorrelation.
chi^2 cal(47.276) > chi^2 tab(6.635) at significance level 1%.

```
. estat bgodfrey, lags(4)
```

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
4	**48.675**	**4**	**0.0000**

H0: no serial correlation

BG test with lags four also rejects the null hypothesis. Presence of autocorrelation.

iii) **Remedial measure-**

A) Resolve Model Misspecification

```
. gen x= gdp* gdp

. reg spending gdp x

      Source |       SS           df       MS      Number of obs   =        62
-------------+----------------------------------   F(2, 59)        =  24531.54
       Model |  15444647.4         2  7722323.69   Prob > F        =    0.0000
    Residual |   18572.704        59  314.791594   R-squared       =    0.9988
-------------+----------------------------------   Adj R-squared   =    0.9988
       Total |  15463220.1        61  253495.411   Root MSE        =    17.742

------------------------------------------------------------------------------
    spending |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |   .5181437   .0099947    51.84   0.000     .4981444    .5381431
           x |   .0000234   3.66e-06     6.40   0.000     .0000161    .0000308
       _cons |   32.61952    3.66923     8.89   0.000     25.27741    39.96163
------------------------------------------------------------------------------

. estat dwatson

Durbin-Watson d-statistic(  3,    62) =  .5268432
```

The value of the d-statistic gets improved but not as much to overcome autocorrelation.

B) First difference

```
.  gen y1=d.spending
(1 missing value generated)

. gen x1=d.gdp
(1 missing value generated)

.  reg y1 x1

      Source |       SS           df       MS      Number of obs   =        61
-------------+----------------------------------   F(1, 59)        =   1774.30
       Model |  193993.512         1  193993.512   Prob > F        =    0.0000
    Residual |  6450.79466        59  109.335503   R-squared       =    0.9678
-------------+----------------------------------   Adj R-squared   =    0.9673
       Total |  200444.307        60  3340.73844   Root MSE        =    10.456

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .5515827   .0130948    42.12   0.000     .5253802    .5777853
       _cons |  -2.097916   1.498842    -1.40   0.167    -5.097093    .9012605
------------------------------------------------------------------------------

. estat dwatson

Durbin-Watson d-statistic(  2,    61) =   1.72268
```

C) Prais-Winsten

The Prais-Winsten estimator considers AR(1) serial correlation of the errors in a linear regression model. The procedure recursively estimates the coefficients and the error autocorrelation of the specified model until sufficient convergence of the AR(1) coefficient is reached.

```
. prais spending gdp,corc

Iteration 0:   rho = 0.0000
Iteration 1:   rho = 0.8736
Iteration 2:   rho = 0.8902
Iteration 3:   rho = 0.8957
Iteration 4:   rho = 0.8977
Iteration 5:   rho = 0.8984
Iteration 6:   rho = 0.8987
Iteration 7:   rho = 0.8988
Iteration 8:   rho = 0.8988
Iteration 9:   rho = 0.8988
Iteration 10:   rho = 0.8989
Iteration 11:   rho = 0.8989
Iteration 12:   rho = 0.8989
```

Cochrane-Orcutt AR(1) regression -- iterated estimates

Source	SS	df	MS
Model	125667.186	1	125667.186
Residual	5222.28274	59	88.5132667
Total	130889.469	60	2181.49115

Number of obs	=	61
F(1, 59)	=	1419.76
Prob > F	=	0.0000
R-squared	=	0.9601
Adj R-squared	=	0.9594
Root MSE	=	9.4081

spending	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp	.5271341	.0139899	37.68	0.000	.4991404	.5551278
_cons	22.64639	12.23766	1.85	0.069	-1.841111	47.1339
rho	.8988575					

Durbin-Watson statistic (original)   0.222004
Durbin-Watson statistic (transformed) 2.076069

The value of dw-statistic changes from 0.222004 to 2.076069. It means the problem of autocorrelation has been resolved.

iv) RESULTS

The problem of autocorrelation has been resolved as the Durbin Watson statistic value is 2.076069. Durbin watson value is greater than 2 and less than 4-dU.