adani | Institute of Digital Technology Management

Time Series Analysis and Forecasting

A Project Report on

Autoregressive Model
Of Groundnut Production In India

Project By:

Aman Verma (20231006)

Harsh Jha (20231021)

Jeet Panchal (20231025)

Snigdha Manocha (20231054)

Vishvarajsinh Chavda (20231059)

Under the supervision of

Dr. Dhyani Mehta

Table of Contents

Abstract

This study inspects the effects of India’s Area under cultivation (Hectare) available for Groundnut production, Yield per Hectare (kg / hectare), pre and post analysis of event on Groundnut production. Using annual time series data from 1985 to 2023, the study used Autoregressive model to scrutinize the relationship. We assume that there is a large autoregressive impact, in which the current year's groundnut output is influenced by the previous year's production, and that the area available for cultivation and yield per hectare are important factors in defining total productivity. Furthermore, we expect specific events identified through pre-post analysis to have discernible impacts on groundnut production, and the autoregressive model will capture these temporal dependencies, providing insights into the complex interplay of factors influencing groundnut farming dynamics.

Statement of Hypothesis

- Null Hypothesis(H_0): The area under cultivation, Yield per Hectare has no significant impact on production of Groundnut.
- Alternate Hypothesis(H_1): The area under cultivation, Yield per Hectare has significant impact on production of Groundnut.

- The first round of tests is conducted to check the normality of all variables-

1. **J-B Test-**

- Null Hypothesis (Ho)- "Data is Normal"
- Alternate Hypothesis(H1)- "Data is not Normal"

2. **Correlogram**

- Null Hypothesis (Ho)- "Data is Normal" (There is no autocorrelation and partial autocorrelation)
- Alternate Hypothesis(H1)- "Data is not Normal" (There is autocorrelation and partial autocorrelation)

3. **Unit root test (ADF)-**

- Null Hypothesis (Ho)- "Data is not normal" (There is unit root)
- Alternate Hypothesis(H1)- "Data is Normal" (There is no unit root)

- **The second round of tests is conducted to check the normality of errors-**

1. **J-B Test-**

- Null Hypothesis (Ho)- “Error is normal”
- Alternate Hypothesis (H1)- “Error is not normal”

2. **Correlogram-**

- Null Hypothesis (Ho)- “Error is normal” (There is no autocorrelation and partial autocorrelation)
- Alternate Hypothesis (H1)- “Error is not normal” (There is autocorrelation and partial autocorrelation)

3. **Unit Root Test (ADF)-**

- Null Hypothesis (Ho)- “Error is not normal” (There is unit root)
- Alternate Hypothesis (H1)- “Error is normal” (There is no unit root)

4. **BG-LM Test-**

- Null Hypothesis (Ho)- “Error is normal” (There is serial correlation)
- Alternate Hypothesis (H1)- “Error is not normal” (There is no serial correlation)

5. **DW-Test-**

- 0 < DW <DL ---> There is (+) autocorrelation
- DL < DW <DU ---> Indecisive zone
- 4-DL < DW < 4 ---> There is (-) autocorrelation
- 4-DU < DW < 4-DL ---> Indecisive zone
- DU < DW < 4-DU ---> No (+) and No (-) autocorrelation

- The third round of tests is conducted to check the impact of event i.e. Event study methodology-

1. **Chow Test-**

- Null Hypothesis (Ho)- “No structural break at specified break points” (No Affect)
- Alternate Hypothesis (H1)- “Structural break at specified break points” (There is Affect)

- The fourth round of tests is conducted to check the heteroscedasticity -

1. **White’s Test-**

- Null Hypothesis (Ho)- “There is no Heteroscedasticity”
- Alternate Hypothesis (H1)- “There is Heteroscedasticity”

2. **Breusch-Pagan-Godfrey Test-**

- Null Hypothesis (Ho)- “There is no Heteroscedasticity”
- Alternate Hypothesis (H1)- “There is Heteroscedasticity”

3. **Glejser Test-**

- Null Hypothesis (Ho)- “There is no Heteroscedasticity”
- Alternate Hypothesis (H1)- “There is Heteroscedasticity”

- **The fifth round of tests is conducted to check the multicollinearity-**
  - VIF (P.Value)<5 ---> "No multicollinearity" (Good)
  - VIF (P.Value)>5 ---> "High multicollinearity" (Bad)

Data Description and Sources

Variable	**Variable Representation**	**Description & Measure**	**Source**
Groundnut	GTLT	Measured: Measured as production of Groundnut in lakh tons. This shows production of groundnut in India measuring in lakh Tons.	RBI
Area	AGTLH	Measured: Measured as Hectare This shows how much area is available for cultivation of groundnut in India.	RBI
Yield	YGTKH	Measured: Measured as Yield (kg / Hectare) This shows how many kg of groundnuts can be produced per hectare.	RBI
Event	Dummy	Absence and presence is denoted with 0 and 1. Here Event is taken as Drought like situation in India in the year 2000.	

Descriptive statistics

	GTLT	AGTLH	YGTKH	DUMMY
Mean	75.36579	64.44737	1211.605	0.605263
Median	74.85000	62.65000	1129.500	1.000000
Maximum	102.8000	87.10000	2063.000	1.000000
Minimum	41.20000	46.00000	694.0000	0.000000
Std. Dev.	16.48438	12.44803	367.3580	0.495355
Skewness	-0.135453	0.272206	0.780671	-0.430706
Kurtosis	2.286163	1.976644	2.743438	1.185507
Jarque-Bera	0.923008	2.127434	3.964053	6.387820
Probability	0.630335	0.345170	0.137790	0.041011
Sum	2863.900	2449.000	46041.00	23.00000
Sum Sq. Dev.	10054.19	5733.275	4993221.	9.078947
Observations	38	38	38	38

The skewness values of all five variables are all negative, except for "AGTLH", which is positive. This means that the distributions of all five variables are skewed to the left, except for "AGTLH", which is skewed to the right.

The kurtosis values of all five variables are all greater than 2, which means that the distributions of all five variables are more peaked than a normal distribution.

The Jarque-Bera statistic is a test of whether the distribution of a variable is normal. The p-values for the Jarque-Bera statistic for all five variables are greater than 0.05, which means that we cannot reject the null hypothesis that the distributions of the variables are normal.

Estimates and Explanation

- **Event study analysis-**

Event	Test	p.value
2000	Chow test	0.0000

In the above table as the p.value<0.05, so we will be accepting "Alternate Hypothesis"' i.e. (H1) which says that structural break at specified break points (In simple words it indicates that there is impact that was observed after the event)

- **Normality Test of Variables-**

Unit root test

Variable	p-value
GTLH	0.4446
AGTLH	0.8642
YGTKH	0.9992

In the above table the p.value of all the variables which includes one "Dependent variable" (GTLH) and 2 "Independent variables" (AGTLH and YGTKH) has p.value>0.05, so we will be accepting "Null Hypothesis" i.e. (Ho) which says that "Data is not normal" and there "Is unit root".

- **Normality test of variable taking log of it-**

Unit root test

Variable	p.value
LOGGT_	0.0000
LOGAGT_	0.0000
LOGYGT_	0.0000

In the above table again, we have calculated the normality of variable because our data turned out to be "Heteroscedastic" in nature, to convert "Heteroscedastic" data in "Homoscedastic" data we calculated the log of each variable. After calculating the log, we will have to calculate the normality of new variables that are calculated using the log.

All the 3 variables in the above table indicates that all the 3 p.value<0.05, so we will be accepting "Alternate Hypothesis" i.e. (H1) which indicates that our "Data is normal" and there is "No unit root".

Dependent Variable: LOGGT_
Method: Least Squares
Date: 01/20/24 Time: 01:36
Sample (adjusted): 1987 2023
Included observations: 37 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000159	0.000523	-0.303194	0.7636
LOGAGT_	1.004496	0.003667	273.9521	0.0000
LOGYGT_	0.998082	0.001264	789.5980	0.0000
DUMMY	0.000382	0.000668	0.572160	0.5711

R-squared	0.999965	Mean dependent var	0.018839
Adjusted R-squared	0.999962	S.D. dependent var	0.318167
S.E. of regression	0.001958	Akaike info criterion	-9.532438
Sum squared resid	0.000126	Schwarz criterion	-9.358284
Log likelihood	180.3501	Hannan-Quinn criter.	-9.471040
F-statistic	316993.2	Durbin-Watson stat	3.013166
Prob(F-statistic)	0.000000		

Equation- LOGGT_ = -0.000159 + 1.004496*LOGAGT_ + 0.998082*LOGYGT_ + 0.000382*DUMMY + ERR

Here the dependent variable is “LOGGT_” and the independent variables are “LOGAGT_”, “LOGYGT_” and “DUMMY”

The p.value of both the independent variables is less than 0.05 i.e. (0.000) which indicates that there will be a significant impact of these 2 variables on the value of dependent variable i.e. “LOGGT_”. Any 1-unit change in value of “LOGAGT_” leads to increase in value of “LOGGT_” by 1.004 times and any 1-unit change in value of “LOGYGT_” leads to increase in value of “LOGGT_” by 0.998 times. Whereas the p.value of “DUMMY” is greater than 0.05 i.e. (0.57) which is insignificant, so it won’t have an impact on the dependent variable i.e. “LOGGT_”.

The value of adjusted R^2 is 0.999 which indicates that our model is good as 99.9% of change in dependent variable due to independent variables is explained by the model rest of the 0.1% which is unexplained by the model is the residual or error.

- **Normality test of error-**

Test	p.value
Jarque bera	0.0000
Correlogram	0.001-0.512
Unit root	0.0001
BG LM	0.0015
DWT	3.013

The 3 major tests that are conducted to check the normality of error among which the superior test is DWT test.

1. **J-B Test-**

According to the table the p.value of J-B test turned out to be less than 0.05 i.e. (0.0000), so we will be accepting "Alternate Hypothesis" i.e. (H1) which indicates that "Error is not normal"

2. **Correlogram Test-**

According to the table the p.value of Correlogram as the majority of the values are greater than 0.05, so we will be accepting "Null Hypothesis" i.e. (Ho) which indicates that "Error is normal"

3. **Unit root Test-**

According to the table the p.value of Unit root turned out to be less than 0.05 i.e. (0.0001), so we will be accepting "Alternate Hypothesis" i.e. (H1) which indicates that "Error is normal" and it "Doesn't have unit root".

4. **BG-LM Test-**

According to the table the p.value of BG-LM turned out to be less than 0.05 i.e. (0.0015), so we will be accepting "Alternate Hypothesis" i.e. (H1) which indicates that "Error is not normal" and "There is serial correlation".

5. **DWT-**

According to the table the DW value i.e. 3.013 which falls between 4 and 4-DL indicates that there is (-) autocorrelation.

- **Checking heteroscedasticity of error-**

Test	p.value
White test	0.8716
Breusch-Pogan-Godfrey test	0.6603
Glejser	0.6519

The three major tests that are conducted to check the heteroscedasticity of the data are mentioned in the above table, among these three all of them are superior test and if any of the single tests says that data is Homoscedastic then we will accept the data as homoscedastic and avoid the rest of the tests of Homoscedasticity. If the residuals become more spread out at higher values in the plot, this is a tell-tale sign that heteroscedasticity is present.

1. **White Test-**

According to the above table the p.value of white test is above 0.05 i.e. (0.87), so we will be accepting “Null Hypothesis” i.e. (Ho), which indicates that “There is no heteroscedasticity”

2. **Breusch-Pogan-Godfrey test-**

According to the above table the p.value of the test is above 0.05 i.e. (0.66), so we will be accepting “Null Hypothesis” i.e. (Ho), which indicates that “There is no heteroscedasticity”

3. **Glejser Test-**

According to the above table the p.value of the test is above 0.05 i.e. (0.65), so we will be accepting “Null Hypothesis” i.e. (Ho), which indicates that “There is no heteroscedasticity”

- **Checking multicollinearity-**

Variance inflation factor	Centred VIF
LOGAGT_	1.1296
LOGYGT_	1.1311
DUMMY	1.0121

According to the above table as all the values of “Centered VIF” <5 which indicates that “There is no multicollinearity” which is good.

Conclusion

Through the application of an autoregressive model, normality testing, heteroscedasticity test, and multicollinearity test, a comprehensive analysis was conducted to understand the complex dynamics influencing groundnut production.

The area under cultivation emerged as a crucial determinant, highlighting the significance of allocating adequate land for groundnut farming. Additionally, the yield per hectare played a pivotal role, underscoring the importance of adopting efficient agricultural practices to enhance productivity. Furthermore, the event study focused on the drought in India in the year 2000 provided valuable insights into the impact of environmental factors on groundnut production.

The autoregressive model facilitated the examination of temporal patterns and dependencies within the data, offering a nuanced understanding of how past conditions influence current outcomes. Normality testing ensured the robustness of the model assumptions, while the heteroscedasticity test addressed potential variations in the variance of the error terms. The multicollinearity test helped identify and mitigate issues arising from interdependencies among the predictor variables.

This comprehensive analysis sheds light on the multifaceted nature of groundnut production, emphasizing the need for a holistic approach that considers cultivation practices, environmental conditions, and historical trends.