

Indian Institute of Technology,

Kharagpur

**Assignment of**

**Econometric Analysis Lab-I**

Topic:

Statistical tests for autocorrelation and

resolving autocorrelation problem

Submitted to: Submitted by:

**Prof. Pulak Mishra Abhishek Tiwari 21HS10004**

**Problem Statement:**

Estimate a regression model to examine how Population growth (annual %) affects by Fertility rate,( total births per woman) of China and test if the estimated model suffers from the problem of autocorrelation. If it is the case, apply some remedial measures to resolve the problem of autocorrelation.

**Model Specification:**

Where is Population growth

is Fertility rate

is random disturbance term that captures the effect of other variables that can affect population growth

**Sources of Data:**

Collected 60 years of continuous time series data from 1965 to 2020 of population growth of China and Fertility rate of China from the Database of **World Bank Open Data**.

<https://data.worldbank.org/>

<https://data.worldbank.org/indicator/SP.POP.GROW?locations=CN>

<https://data.worldbank.org/indicator/SP.DYN.TFRT.IN>

**Measurement of the Variables:**

* **Population growth (annual %)**

Population growth is the increase in the number of people in a population  or dispersed group. 

The "population growth rate" is the rate at which the number of individuals in a population increases in a given time period, expressed as a fraction of the initial population. Specifically, population growth rate refers to the change in population over a unit time period, often expressed as a percentage of the number of individuals in the population at the beginning of that period. This can be written as the formula, valid for a sufficiently small time interval:



* **Fertility rate, total (births per woman)**

The average number of children a hypothetical cohort of women would have at the end of their reproductive period if they were subject during their whole lives to the fertility rates of a given period and if they were not subject to mortality. It is expressed as children per woman.

Total fertility rate is directly calculated as the sum of age-specific fertility rates (usually referring to women aged 15 to 49 years), or five times the sum if data are given in five-year age groups. An age- or age-group-specific fertility rate is calculated as the ratio of annual births to women at a given age or age-group to the population of women at the same age or age-group, in the same year, for a given country, territory, or geographic area. Population data from the United Nations correspond to mid-year estimated values, obtained by linear interpolation from the corresponding United Nations fertility medium-variant quinquennial population projections

**Estimation:**

Model :

First we check significance of model thorugh regression results

Then we apply runs test , DW test and BG test to check autocorrelation.

And then we go for remedies either through first difference method or through General Difference Method

* + - * 1. **Regression Results:**

****

The estimated model is significant at 1% level of significance.

Fertility rate is statistically significant too at 5% level of

significance.

* + - * 1. **Tests for autocorrelation:**

1. Plotting Scatter plot of the residuals and :

We can clearly observe that and are positively related, we suspect that there is a possibility of **Positive Autocorrelation.**

1. **Runs Test:**

The number of runs (R) follows asymptotically normal distribution when n1 > 10 and n2 > 10

: = 0 (No autocorrelation)

: Presence of autocorrelation)



E(Runs) = + 1 = 29

Var (Runs) = = 13.74545

SD (Runs) = 3.707486

Lower limit = E(R)-1.96\*SD = 36.26667

Upper limit = E(R)+1.96\*SD = 21.73333

Since N(runs)(5)<lower limit(36.2667)

Hence Null Hypothesis is rejected that shows the presence of **Positive Autocorrelation.**

1. **Durbin-Watson D test:**

**Assumption:**

follows first order autocorrelation structure i.e., where follows all the assumption of a random disturbance term in regression analysis.

**Null Hypothesis:**

: = 0 (No autocorrelation)

Presence of autocorrelation)

**More assumption on :**

E () = 0 E ( ) =0 E ( =0

E ( =

d= = 0.2078127

(1-d/2) **0.896094**

for n=56 and k=1 at 1% significance level

Dl = 1.36 Du = 1.43

Now, as d<Dl i.e., signifies presence of **Positive autocorrelation**

1. **Breusch Godfrey Test:**

Here we can relax the assumption of first order autocorrelation structure and take any higher order autocorrelation structure

Now, estimating the model:

Testing the Null Hypothesis : (No autocorrelation)



Null Hypothesis is Rejected, signifies **presence of autocorrelation**

**Runs Test , Durbin-Watson d Test** and **Breusch Godfrey Test** signifies presence of positive autocorrelation

* + - * 1. **Remedial measures for problem of autocorrelation:**

1. **First-Difference Model: If the autocorrelation coefficient () is not significantly different from 1 (one)**

**Barenblat Webb Test:**

Regress the first difference model i.e., where =

Test Statistics: for testing the null hypothesis : = 1

For our data on computing we get  **= 0.170008**

Since Dl from Durbin Watson Table is 1.36 (at 1% significance level for n=55 and k=1) i.e**.,** our **is less than Dl**.

Hence null hypothesis is not rejected and is not significantly different from 1.

Now, we can proceed with **first difference model** to solve the problem of autocorrelation.

**First Difference Model:**

Regress the first difference model i.e.,

where =

**Regression Results:**



The estimated model is significant at 1% level of significance.

Fertility rate is statistically significant too at 5% level of

significance.

Plotting Scatter plot of the residuals and :

We can observe that residuals and are very much **scattered**, we suspect there is **no presence of Autocorrelation.**

**Runs Test:**

: = 0 (No autocorrelation)

Presence of autocorrelation)



E(Runs) = + 1 = 28.49091

Var (Runs) = = 13.48628

SD (Runs) = 3.672367

Lower limit = E(R)-1.96\*SD = 21.29307

Upper limit = E(R)+1.96\*SD = 35.68875

Since N(runs) is in between lower and upper limit. Hence Null Hypothesis is not rejected and there is **No presence of Autocorrelation.**

**Durbin-Watson D test:**

d= = 1.381182

(1-d/2) 0.082

Since, for n=55 and k=1 at 1% significance level, value of Dl and Du from Durbin-Watson table are

Dl = 1.36 Du = 1.43

Now, as Dl<d<Du i.e., Value of the DW d statistic lies in the zone of **indecision**. So we can’t conclude anything from DW D test

**Breusch Godfrey Test:**

Estimating the model:

Testing the Null Hypothesis : (No autocorrelation)

For n=59, k=4 i.e., taking 4 lags, we got the value of = 0.049

~



Null Hypothesis is not Rejected, signifies **absence of autocorrelation**

**Results and Discussions:**

* We first tested the problem of autocorrelation in the original model i.e., Yt = α + βXt + ut and got Positive autocorrelation using all three tests for autocorrelation.
* After applying Barenblat Webb test, we have applied first difference model and again checked for problem of autocorrelation. Applying first difference model, the problem of autocorrelation got resolved and it can be cleared from plotting vt and vt−1. No Autocorrelation problem in first difference model.
* We can also have problem of specific error and autoregressive problems but checking for those problems are beyond the scope of assignment.

**Conclusion:**

Our analysis revealed evidence of autocorrelation in the time series data of population growth and fertility rate. The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots showed significant lags beyond the confidence interval, indicating that the current values of the variables were significantly correlated with their past values. To correct for the autocorrelation, we used the first differential method, which involved taking the first difference of the time series data. After applying the first differential method, we re-checked the through plot, runs test , dw d test and BG test

Based on the evidence of autocorrelation in the original data and the success of the first differential method in correcting for it, we conclude that the assumptions of independence of observations were initially violated but were addressed through the use of the first differential method. Therefore, we recommend using the first differential method to account for the autocorrelation in the data and obtain accurate predictions. However, it is important to note that the first differential method may not always be the most appropriate or effective method for correcting for autocorrelation like Generalized Difference Method.