PANJAB UNIVERSITY,CHANDIGARH

Department of Statistics

Academic Year – 2021 – 22

Time Series Forecasting of GDP in India using ARIMA Model

Submitted By:

Diksha

Gurbinder Singh

Manisha Rani

Shikha

Acknowledgement

With colossal pleasure , we are presenting "**Time series forecasting of GDP in India using ARIMA model"** report as part of the curriculum of **' Master of Statistics'** .

We are thankful to the **Department of Statistics, Panjab University Chandigarh** for giving us the opportunity to work on this project. We would like to express our special thanks to our mentor, **Dr. Gulfam , Assistant Professor in Statistics ,NCERT,** for their valuable guidance, keen interest and encouragement throughout. We are grateful to all the other teachers for their kind co-operation and support. A special thanks to our team mates who proactively worked to complete the project on time. Thanks to all the other groups for creating such a competitive environment that encouraged us to improve our project consistently. It has been a great learning experience.

Abstract

Gross Domestic Product is one of the most important economic indicators of the country and its positive or negative growth indicates the economic development of the country. It is calculated quarterly and yearly at the end of the financial year. The GDP growth of India has seen fluctuations from last few decades after independence and reached as high as 10.25 in 2010 and declined to low of -5.23 in 1979. The GDP growth has witnessed a continuous decline in the past five years, taking it from 8.15 in 2015 to 1.87 in 2020.The lockdown imposed in the country to curb the spread of COVID-19 has caused massive slowdown in the economy of the country by affecting all major contributing sectors of the GDP except agricultural sector. To keep on track on the GDP growth is one of the parameters for deciding the economic policies of the country. In this study, we are analyzing and forecasting the GDP growth using the time series forecasting techniques Prophet and Arima model. This model can assist policy makers in framing policies or making decisions.

OBJECTIVES

- Analysis of GDP growth trend in India from 1951 to 2020 (Annually data)
- Forecasting of GDP growth trend in India from 2021 to 2030
- Analysis of GDP growth trend in India from 2005 to 2020 (Quarterly data)
- Forecasting of GDP growth trend in India from 2021 to 2024

Contents

1) INTRODUCTION

1.1) TIME SERIES:

WHAT IS TIME SERIES?

A time series is a collection of observations of well-defined data items obtained through repeated measurements over time. For example, measuring the value of retail sales each month of the year would comprise a time series. This is because sales revenue is well defined, and consistently measured at equally spaced intervals. Data collected irregularly or only once are not time series.

Time series data is everywhere, since time is a constituent of everything that is observable. As our world gets increasingly instrumented, sensors and systems are constantly emitting a relentless stream of time series data. Such data has numerous applications across various industries. Let’s put this in context through some examples.

Examples of time series analysis:

1.Electrical activity in the brain

2.Rainfall measurements

3.Stock prices

4.Annual retail sales

5.Monthly subscribers

6.Heartbeats per minute

An observed time series can be decomposed into three components: the trend (long term direction), the seasonal (systematic, calendar related movements) and the irregular (unsystematic, short term fluctuations).

WHAT IS SEASONALITY?

The seasonal component consists of effects that are reasonably stable with respect to timing, direction and magnitude. It arises from systematic, calendar related influences such as:

- **Natural Conditions**

  weather fluctuations that are representative of the season (uncharacteristic weather patterns such as snow in summer would be considered irregular influences)
- **Business and Administrative procedures**

  start and end of the school term
- **Social and Cultural behaviour**

  Christmas

It also includes calendar related systematic effects that are not stable in their annual timing or are caused by variations in the calendar from year to year, such as:

- <u>**Trading Day Effects**</u>

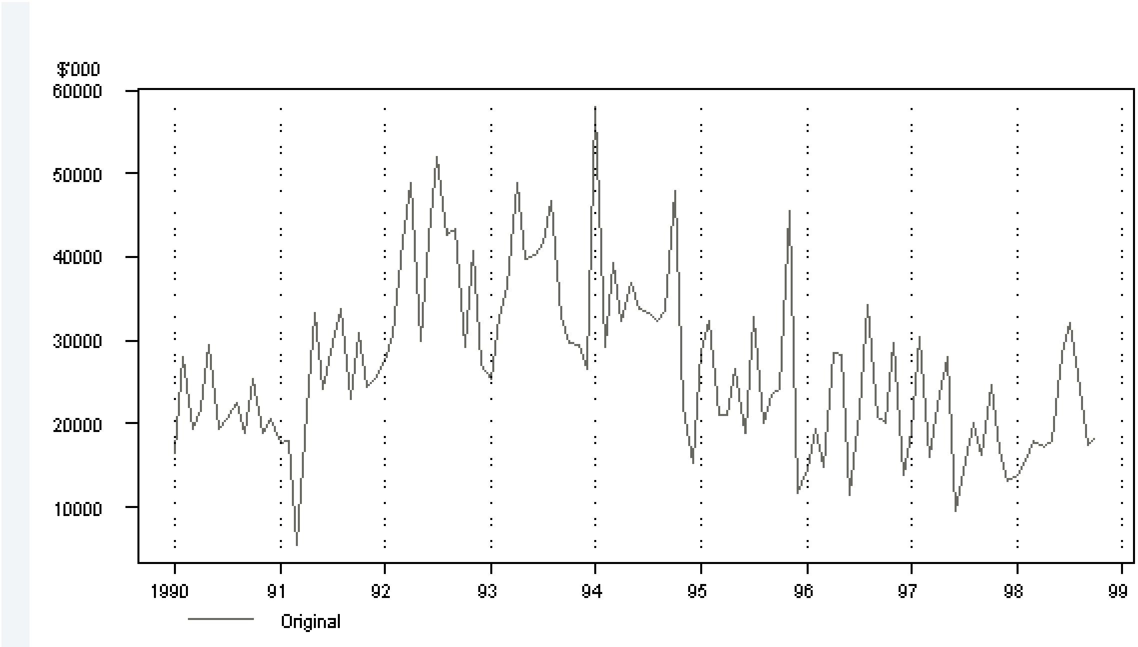
  the number of occurrences of each of the day of the week in a given month will differ from year to year
  - There were 4 weekends in March in 2000, but 5 weekends in March of 2002
- <u>**Moving Holiday Effects**</u>

  holidays which occur each year, but whose exact timing shifts
  - Easter, Chinese New Year

HOW DO WE IDENTIFY SEASONALITY?

Seasonality in a time series can be identified by regularly spaced peaks and troughs which have a consistent direction and approximately the same magnitude every year, relative to the trend. The following diagram depicts a strongly seasonal series. There is an obvious large seasonal increase in December retail sales in New South Wales due to Christmas shopping. In this example, the magnitude of the seasonal component increases over time, as does the trend.

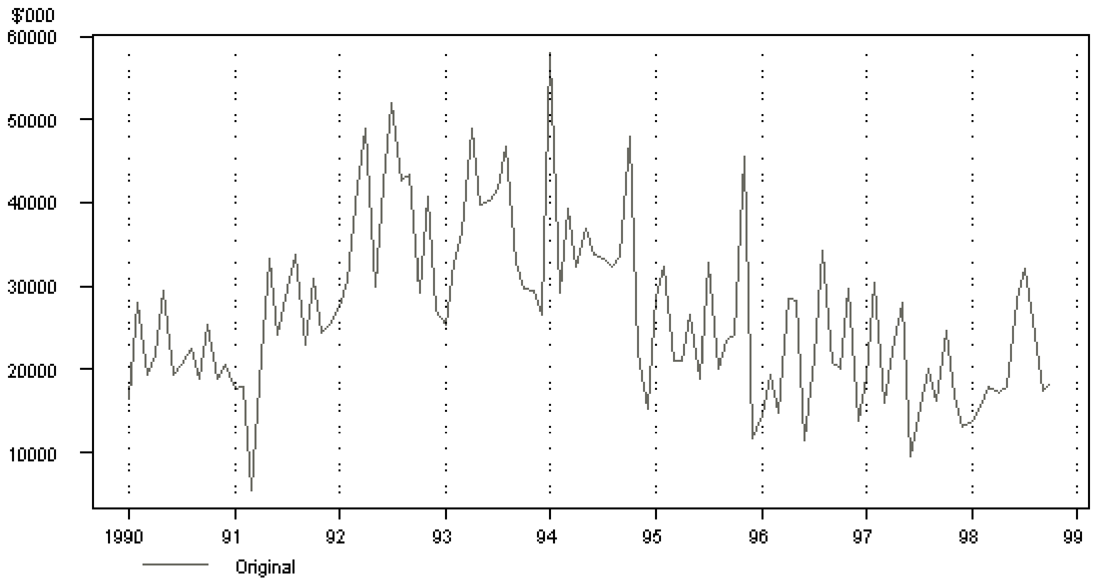
Figure 1: Monthly Retail Sales in New South Wales (NSW) Retail Department Stores



WHAT IS AN IRREGULAR?

The irregular component (sometimes also known as the residual) is what remains after the seasonal and trend components of a time series have been estimated and removed. It results from short term fluctuations in the series which are neither systematic nor predictable. In a highly irregular series, these fluctuations can dominate movements, which will mask the trend and seasonality. The following graph is of a highly irregular time series:

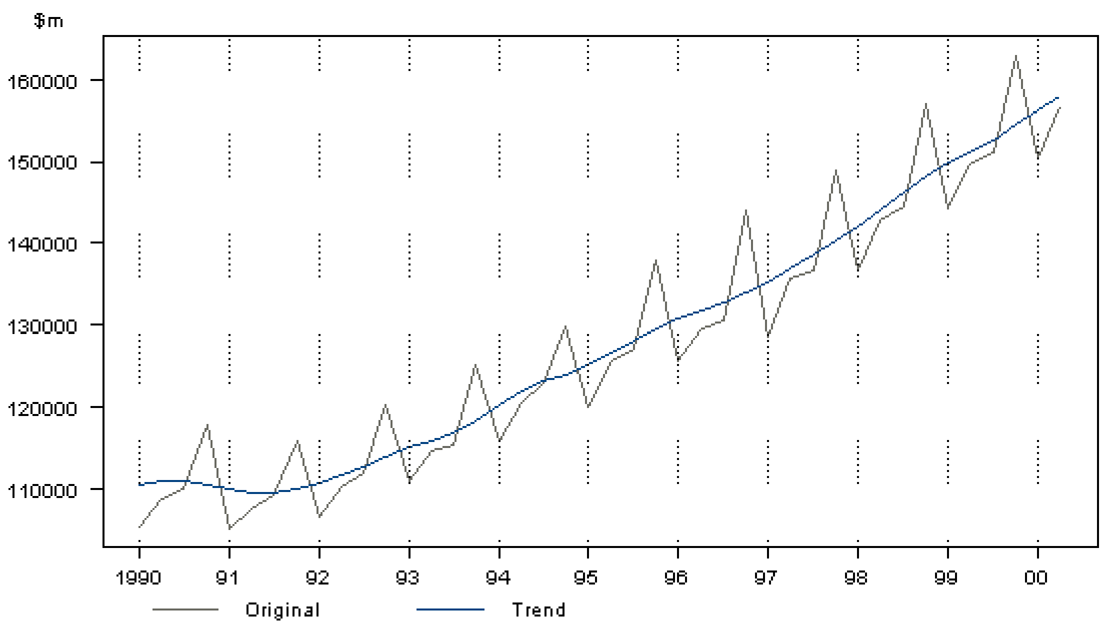
Figure 2: Monthly Value of Building Approvals, Australian Capital Territory (ACT)



WHAT IS THE TREND?

The trend is defined as the 'long term' movement in a time series without calendar related and irregular effects, and is a reflection of the underlying level. It is the result of influences such as population growth, price inflation and general economic changes. The following graph depicts a series in which there is an obvious upward trend over time:

FIGURE2:QUARTERLY GROSS DOMESTIC PRODUCT



WHAT ARE THE UNDERLYING MODELS USED TO DECOMPOSE THE OBSERVED TIME SERIES?

Decomposition models are typically additive or multiplicative, but can also take other forms such as pseudo-additive.

Additive Decomposition

In some time series, the amplitude of both the seasonal and irregular variations do not change as the level of the trend rises or falls. In such cases, an additive model is appropriate.

In the additive model, the observed time series (O_t) is considered to be the sum of three independent components: the seasonal S_t, the trend T_t and the irregular
I_t.

$$\text{Observed series} = \text{Trend} + \text{Seasonal} + \text{Irregular}$$

That is

$$O_t = T_t + S_t + I_t$$

Each of the three components has the same units as the original series. The seasonally adjusted series is obtained by estimating and removing the seasonal effects from the original time series. The estimated seasonal component is denoted by $\hat{S}_t$ The seasonally adjusted estimates can be expressed by:

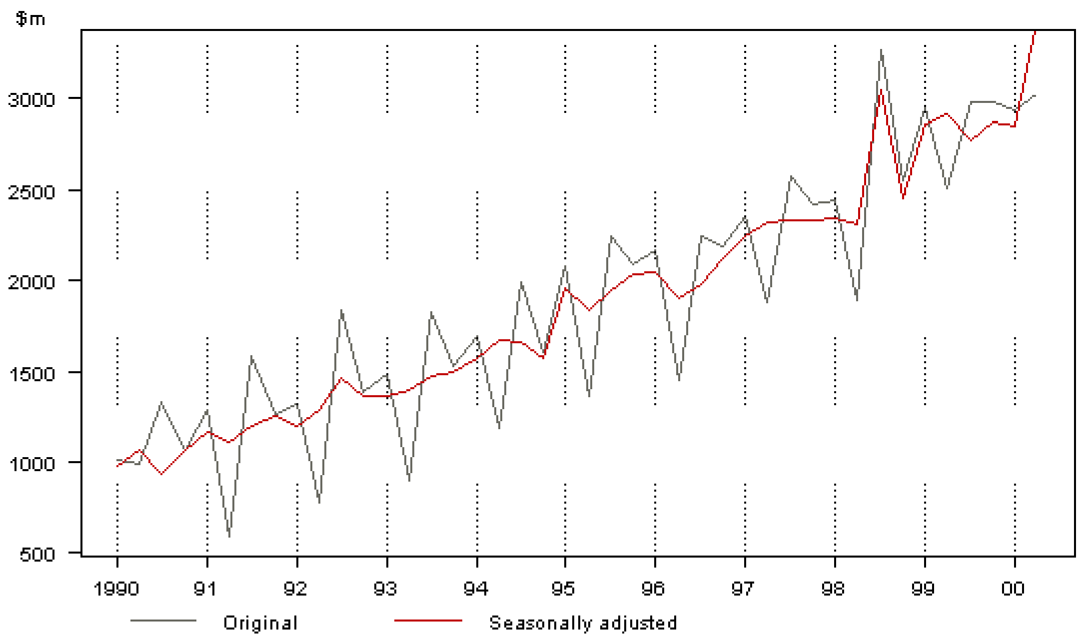
$$\begin{aligned}\text{Seasonally adjusted series} &= \text{Observed series} - \text{Seasonal} \\ &= \text{Trend} + \text{Irregular}\end{aligned}$$

In symbols,

$$\begin{aligned}SA_t &= O_t - \hat{S}_t \\ &= T_t + I_t\end{aligned}$$

The following figure depicts a typically additive series. The underlying level of the series fluctuates but the magnitude of the seasonal spikes remains approximately stable.

Figure 4: General Government and Other Current Transfers to Other Sectors



Multiplicative Decomposition

In many time series, the amplitude of both the seasonal and irregular variations increase as the level of the trend rises. In this situation, a multiplicative model is usually appropriate.

In the multiplicative model, the original time series is expressed as the product of trend, seasonal and irregular components.

$$\text{Observed series} = \text{Trend} \times \text{Seasonal} \times \text{Irregular}$$

or

$$O_t = T_t \times S_t \times I_t$$

The seasonally adjusted data then becomes:

$$\begin{aligned}\text{Seasonally Adjusted series} &= \text{Observed} \div \text{Seasonal} \\ &= \text{Trend} \times \text{Irregular}\end{aligned}$$

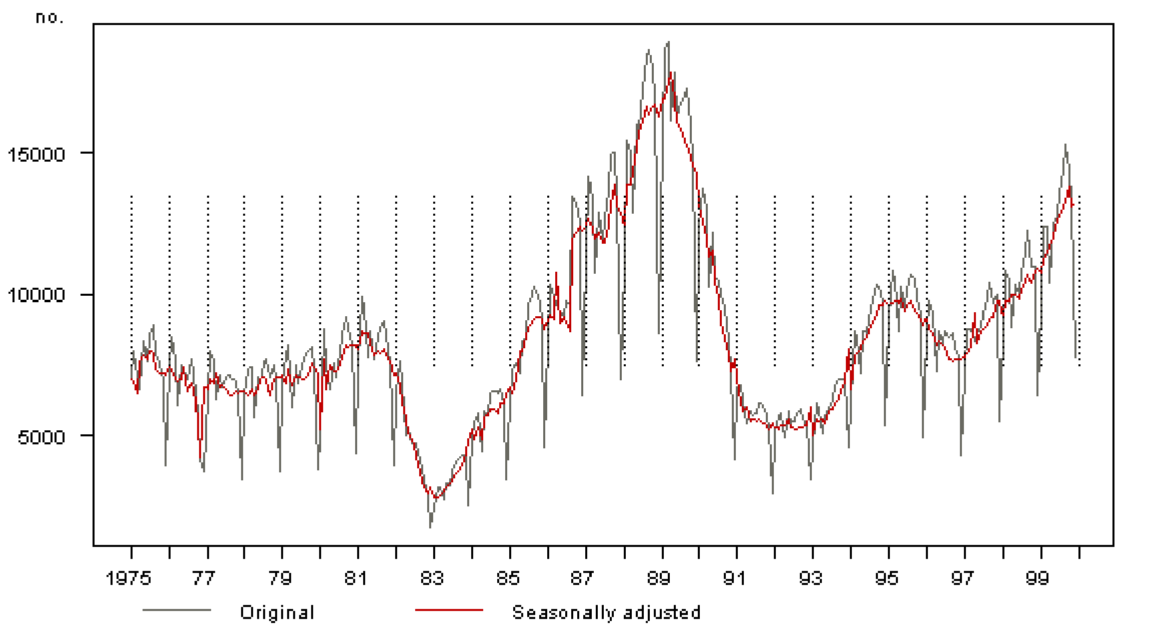
or

$$SA_t = O_t / \hat{S}_t = T_t \times I_t$$

Under this model, the trend has the same units as the original series, but the seasonal and irregular components are unitless factors, distributed around 1.

Most of the series analysed by the ABS show characteristics of a multiplicative model. As the underlying level of the series changes, the magnitude of the seasonal fluctuations varies as well.

Figure 5: Monthly NSW ANZ Job Advertisements



1.2) ABOUT TIME SERIES ANALYSIS:

Time series modeling is a dynamic research area which has attracted attentions of researchers community over last few decades. The main aim of time series modeling is to carefully collect and rigorously study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series. This model is then used to generate future values for the series, i.e. to make forecasts. Time series forecasting thus can be termed as the act of predicting the future by understanding the past. Due to the indispensable importance of time series forecasting in numerous practical fields such as business, economics, finance, science and engineering, etc., proper care should be taken to fit an adequate model to the underlying time series. It is obvious that a successful time series forecasting depends on an appropriate model fitting. A lot of efforts have been done by researchers over many years for the development of efficient models to improve the forecasting accuracy. As a result, various important time series forecasting models have been evolved in literature. One of the most popular and frequently used stochastic time series models is the Autoregressive Integrated Moving Average (ARIMA) model. The basic assumption made to implement this model is that the considered time series is linear and follows a particular known statistical distribution, such as the normal distribution. ARIMA model has subclasses of other models, such as the Autoregressive (AR) , Moving Average (MA) and Autoregressive Moving Average (ARMA) models. For seasonal time series forecasting, Box and Jenkins had proposed a quite successful variation of ARIMA model, viz. the Seasonal ARIMA (SARIMA) .The popularity of the ARIMA model is mainly due to its flexibility to represent several varieties of time series with simplicity as well as the associated Box-Jenkins methodology for optimal model building process.

WHAT IS STATIONARITY ?

The concept of stationarity of a stochastic process can be visualized as a form of statistical equilibrium .The statistical properties such as mean and variance of a stationary process do not depend upon time. It is a necessary condition for building a time series model that is useful for future forecasting. Further, the mathematical complexity of the fitted model reduces with this assumption. There are two types of stationary processes which are defined below:

1.3) SELECTION OF MODEL:

The selection of a proper model is extremely important as it reflects the underlying structure of the series and this fitted model in turn is used for future forecasting. A time series model is said to be linear or non-linear depending on whether the current value of the series is a linear or non-linear function of past observations. In general models for time series data can have many forms and represent different stochastic processes. There are two widely used linear time series models in literature, viz. Autoregressive (AR) and Moving Average (MA) models. Combining these two, the Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average

(ARIMA) models have been proposed in literature. The Autoregressive Fractionally Integrated Moving Average (ARFIMA) model generalizes ARMA and ARIMA models. For seasonal time series forecasting, a variation of ARIMA, viz. the Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used. ARIMA model and its different variations are based on the famous Box-Jenkins principle and so these are also broadly known as the Box-Jenkins models.

THE AUTOREGRESSIVE MOVING AVERAGE (ARMA) MODEL:

An ARMA(p, q) model is a combination of AR(p) and MA(q) models and is suitable for univariate time series modeling. In an AR(p) model the future value of a variable is assumed to be a linear combination of p past observations and a random error together with a constant term. Usually For estimating parameters of an AR process using the given time series, the YuleWalker equations are used. Just as an AR(p) model regress against past values of the series, an MA(q) model uses past errors as the explanatory variables.The random shocks are assumed to be a white noise process, i.e. a sequence of independent and identically distributed random variables with zero mean and a constant variance .Generally, the random shocks are assumed to follow the typical normal distribution. Thus conceptually a moving average model is a linear regression of the current observation of the time series against the random shocks of one or more prior observations. Fitting an MA model to a time series is more complicated than fitting an AR model because in the former one the random error terms are not fore-seeable.

Autoregressive (AR) and moving average (MA) models can be effectively combined together to form a general and useful class of time series models, known as the ARMA models. Here the model orders qp, refer to p autoregressive and q moving average terms. Usually ARMA models 8 are manipulated using the lag operator notation.Polynomials of lag operator or lag polynomials are used to represent ARMA models. Stationarity Analysis: It is proved by Box and Jenkins that a necessary and sufficient condition for the AR(p) process to be stationary is that all the roots of the characteristic equation must fall outside the unit circle. Hipel and McLeod mentioned another simple algorithm (by Schur and Pagano) for determining stationarity of an AR process.AnMA(q) process is always stationary, irrespective of the values the MA parameters . The conditions regarding stationarity and invertibility of AR and MA processes also hold for an ARMA process. An ARMA(p, q) process is stationary if all the roots of the characteristic equation lie outside the unit circle. Similarly, if all the roots of the lag equation lie outside the unit circle, then the ARMA(p, q) process is invertible and can be expressed as a pure AR process.

Autoregressive Integrated Moving Average (ARIMA):

Models The ARMA models, described above can only be used for stationary time series data. However in practice many time series such as those related to socio-economic business show non-stationary behavior. Time series, which contain trend and seasonal patterns, are also non-stationary in nature.Thus from application view point , ARMA models are inadequate to properly describe non-stationary time series, which are frequently encountered in practice. For this reason the ARIMA model is 9 proposed, which is a generalization of an ARMA model to include the case of non-stationarity as well. In ARIMA models a non-stationary time series is made stationary by applying finite differencing of the data points.AnARIMA(p,0,0) is nothing but the AR(p) model and ARIMA(0,0,q) is the MA(q) model. ARIMA(0,1,0), is a special one and known as the Random Walk model .It is widely used for non-stationary data, like economic and stock price series. A useful generalization of ARIMA models is the Autoregressive Fractionally Integrated Moving Average (ARFIMA) model, which allows non-integer values of the differencing parameter d. ARFIMA has useful application in modeling time series with long memory. In this model the expansion of the term is to be done by using the general binomial theorem. Various contributions have been made by researchers towards the estimation of the general ARFIMA parameters.

Seasonal ARIMA (SARIMA):

Box and Jenkins have generalized this model to deal with seasonality. Their proposed model is known as the Seasonal ARIMA (SARIMA) model. In this model seasonal differencing of appropriate order is used to remove non-stationarity from the series. A first order seasonal difference is the difference between an observation and the corresponding observation from the previous year.

1.4) Gross Domestic Product (GDP)

What Is Gross Domestic Product (GDP)?

Gross domestic product (GDP) is the total monetary or market value of all the finished goods and services produced within a country's borders in a specific time period. As a broad measure of overall domestic production, it functions as a comprehensive scorecard of a given country's economic health.

Though GDP is typically calculated on an annual basis, it is sometimes calculated on a quarterly basis as well. In the U.S., for example, the government releases an annualized GDP estimate for each fiscal quarter and also for the calendar year. The individual data sets included in this report are given in real terms, so the data is adjusted for price changes and is, therefore, net of inflation.

History of GDP

The concept of GDP was first proposed in 1937 in a report to the U.S. Congress in response to the Great Depression, conceived of and presented by an economist at the National Bureau of Economic Research, Simon Kuznets.[8]

Beginning in the 1950s, however, some economists and policy-makers began to question GDP. Some observed, for example, a tendency to accept GDP as an absolute indicator of a nation's failure or success, despite its failure to account for health, happiness, (in)equality, and other constituent factors of public welfare. In other words, these critics drew attention to a distinction between economic progress and social progress.

However, most authorities, like Arthur Okun, an economist for President John F. Kennedy's Council of Economic Advisers, held firm to the belief that GDP is an absolute indicator of economic success, claiming that for every increase in GDP, there would be a corresponding drop in unemployment.[9]

KEY TAKEAWAYS

- Gross domestic product (GDP) is the monetary value of all finished goods and services made within a country during a specific period.

- GDP provides an economic snapshot of a country, used to estimate the size of an economy and growth rate.
- GDP can be calculated in three ways, using expenditures, production, or incomes. It can be adjusted for inflation and population to provide deeper insights.
- Though it has limitations, GDP is a key tool to guide policy-makers, investors, and businesses in strategic decision-making.

Types of Gross Domestic Product

GDP can be reported in several ways, each of which provides slightly different information.

Nominal GDP

Nominal GDP is an assessment of economic production in an economy that includes current prices in its calculation. In other words, it doesn't strip out inflation or the pace of rising prices, which can inflate the growth figure.

All goods and services counted in nominal GDP are valued at the prices that those goods and services are actually sold for in that year. Nominal GDP is evaluated in either the local currency or U.S. dollars at currency market exchange rates to compare countries' GDPs in purely financial terms.

Nominal GDP is used when comparing different quarters of output within the same year. When comparing the GDP of two or more years, real GDP is used. This is because, in effect, the removal of the influence of inflation allows the comparison of the different years to focus solely on volume.

Real GDP

Real GDP is an inflation-adjusted measure that reflects the quantity of goods and services produced by an economy in a given year, with prices held constant from year to year to separate out the impact of inflation or deflation from the trend in output over time. Since GDP is based on the monetary value of goods and services, it is subject to inflation.

Rising prices will tend to increase a country's GDP, but this does not necessarily reflect any change in the quantity or quality of goods and services produced. Thus, by looking just at an economy's nominal GDP, it can be difficult to tell whether the figure has risen because of a real expansion in production or simply because prices rose.

Real GDP is calculated using a GDP price deflator, which is the difference in prices between the current year and the base year. For example, if prices rose by 5% since the base year, then the deflator would be 1.05. Nominal GDP is divided by this deflator, yielding real GDP. Nominal GDP is usually higher than real GDP because inflation is typically a positive number.

GDP Per Capita

GDP per capita is a measurement of the GDP per person in a country's population. It indicates that the amount of output or income per person in an economy can indicate average productivity or average living standards. GDP per capita can be stated in nominal, real (inflation-adjusted), or PPP (purchasing power parity) terms.

At a basic interpretation, per-capita GDP shows how much economic production value can be attributed to each individual citizen. This also translates to a measure of overall national wealth since GDP market value per person also readily serves as a prosperity measure.

GDP Growth Rate

The GDP growth rate compares the year-over-year (or quarterly) change in a country's economic output to measure how fast an economy is growing. Usually expressed as a percentage rate, this measure is popular for economic policy-makers because GDP growth is thought to be closely connected to key policy targets such as inflation and unemployment rates.

If GDP growth rates accelerate, it may be a signal that the economy is "overheating" and the central bank may seek to raise interest rates. Conversely, central banks see a shrinking (or negative) GDP growth rate (i.e., a recession) as a signal that rates should be lowered and that stimulus may be necessary.

GDP Purchasing Power Parity (PPP)

While not directly a measure of GDP, economists look at purchasing power parity (PPP) to see how one country's GDP measures up in "international dollars" using a method that adjusts for differences in local prices and costs of living to make cross-country comparisons of real output, real income, and living standards.

How to Calculate GDP?

GDP can be calculated using three primary methods. All three methods should produce the same figures when calculated separately. These three approaches are known as the expenditure approach, the output (or production) approach, and the income approach.

GDP Calculation Approaches

---

1. The Expenditure Approach

   The expenditure approach, also known as the spending approach, firstly calculates the spending carried out by the different groups that actively participate in the economy. The Indian GDP is primarily measured based on the expenditure approach.
   The GDP can be calculated using this approach by the following formula:

$$GDP = C + G + I + NX$$

   where C=consumption; G=government spending; I=Investment; and NX=net exports, all these activities are attributed to the GDP of a country.

- Consumption in the above formula refers to consumer spending or private consumption expenditure. Consumers spend money to monopolize goods and services, such as groceries and beauty treatment. Consumer spending is the biggest contributing component of GDP, accounting for more than two-thirds of the Indian GDP. Consumer confidence, therefore, has a very significant effect on economic growth.

On a scale a high confidence level indicates that consumers are willing to splurge money on goods and services, while a low confidence level reflects reluctancy about the future and an unwillingness to spend.

- Investment refers to capital expenditures or private domestic investments. Businesses need to spend money in order to invest in their business activities and hence grow their business.

  Business investment critically contributes to a nation's GDP since it helps in increasing the productive capacity of an economy and also generate more employment opportunities.

- Net exports is the calculation that involves subtracting total exports taken place from total imports taken place (NX = Exports - Imports).

  The goods and services that an economy produces and is exported to other countries, less the number of imports that are purchased by domestic consumers of an economy, represents a country's net exports.

  All expenditures used by companies located in a certain country, even if the companies are foreign, are included in this calculation.

2. <u>The Production (Output) Approach</u>

The production approach is actually the reverse of the expenditure approach as it measures the input costs that contribute to economic activity while the production approach estimates the total value of economic output and deducts the cost of goods that are consumed during the process (like those of materials and services) from it.

Whereas the expenditure approach extends forward from costs, the production approach retracts steps from the point of a state of completed economic activity.

3. <u>The Income Approach</u>

The income approach represents a kind of middle landmark between the above two approaches for calculating GDP.

- The income approach involves calculating the income earned by all the leading factors of production in an economy, including the wages received by labor, the rent earned by land mortgaging, the return on capital in the form of interest, and corporate profits.
- The income approach tends to make some adjustments for those items that do not contribute as payments made to factors of production. For an instance, there are some taxes—such as sales taxes and property taxes—that are regarded as indirect business taxes.

- In addition, depreciation–a buffer that businesses keep aside to account for the replacement of equipment that tends to wear down due to rigorous usage, is also added to the national income. All of these payments and finances together constitutes a given nation's income.

GDP vs GNP vs GNI

Although GDP is a widely used metric, there are other ways of measuring the economic growth of a country. While GDP measures the economic activity within the physical borders of a country (whether the producers are native to that country or foreign-owned entities), gross national product (GNP) is a measurement of the overall production of people or corporations native to a country, including those based abroad. GNP excludes domestic production by foreigners.

Gross national income (GNI) is another measure of economic growth. It is the sum of all income earned by citizens or nationals of a country (regardless of whether the underlying economic activity takes place domestically or abroad). The relationship between GNP and GNI is similar to the relationship between the production (output) approach and the income approach used to calculate GDP.

GNP uses the production approach, while GNI uses the income approach. With GNI, the income of a country is calculated as its domestic income, plus its indirect business taxes and depreciation (as well as its net foreign factor income). The figure for net foreign factor income is calculated by subtracting all payments made to foreign companies and individuals from all payments made to domestic businesses.

In an increasingly global economy, GNI has been put forward as a potentially better metric for overall economic health than GDP. Because certain countries have most of their income withdrawn abroad by foreign corporations and individuals, their GDP figure is much higher than the figure that represents their GNI.

For example, in 2019, Luxembourg had a significant difference between its GDP and GNI, mainly due to large payments made to the rest of the world via foreign corporations that did business in Luxembourg, attracted by the tiny nation's favorable tax laws.[4] On the contrary, in the U.S., GNI and GDP do not differ substantially. In 2019, U.S. GDP was $21.7 trillion while its GNI was $21.7 trillion also.[56]

Sources for GDP Data

The World Bank hosts one of the most reliable web-based databases. It has one of the best and most comprehensive lists of countries for which it tracks GDP data. The International Money Fund (IMF) also provides GDP data through its multiple databases, such as World Economic Outlook and International Financial Statistics.

Another highly reliable source of GDP data is the Organization for Economic Cooperation and Development (OECD). The OECD not only provides historical data but also forecasts GDP growth. The disadvantage of using the OECD database is that it tracks only OECD member countries and a few nonmember countries.

In the U.S., the Fed collects data from multiple sources, including a country's statistical agencies and The World Bank. The only drawback to using a Fed database is a lack of updating in GDP data and an absence of data for certain countries.

The Bureau of Economic Analysis (BEA) a division of the U.S. Department of Commerce, issues its own analysis document with each GDP release, which is a great investor tool for analyzing figures and trends and reading highlights of the very lengthy full release.

How to Use GDP Data

Most nations release GDP data every month and quarter. In the U.S., the Bureau of Economic Analysis (BEA) publishes an advance release of quarterly GDP four weeks after the quarter ends, and a final release three months after the quarter ends. The BEA releases are exhaustive and contain a wealth of detail, enabling economists and investors to obtain information and insights on various aspects of the economy.[1]

GDP's market impact is generally limited, since it is "backward-looking," and a substantial amount of time has already elapsed between the quarter-end and GDP data release. However, GDP data can have an impact on markets if the actual numbers differ considerably from expectations.

Because GDP provides a direct indication of the health and growth of the economy, businesses can use GDP as a guide to their business strategy. Government entities, such as the Fed in the U.S., use the growth rate and other GDP stats as part of their decision process in determining what type of monetary policies to implement.

If the growth rate is slowing, they might implement an expansionary monetary policy to try to boost the economy. If the growth rate is robust, they might use monetary policy to slow things down to try to ward off inflation.

Real GDP is the indicator that says the most about the health of the economy. It is widely followed and discussed by economists, analysts, investors, and policy-makers. The advance release of the latest data will almost always move markets, although that impact can be limited, as noted above.

Which Country Has the Highest GDP?

The countries with the two highest GDPs in the world are the United States and China. However, their ranking differs depending on how you measure GDP. Using nominal GDP, the United States comes in first with a GDP of $20.89 trillion as of 2020, compared to $14.7 trillion for China.[12][13] Many economists, however, argue that it is more accurate to use purchasing power parity (PPP) GDP as a measure for national wealth. By this metric, China is actually the world leader with a 2020 PPP GDP of $24.3 trillion, followed by $20.9 trillion for the United States.[14][15]

Is a High GDP Good?

Most people perceive a higher GDP to be a good thing because it is associated with greater economic opportunities and an improved standard of material well-being. It is possible, however, for a country to have a high GDP and still be an unattractive place to live, so it is important to also consider other measurements. For example, a country could have a high GDP and a low per-capita GDP, suggesting that significant wealth exists but is concentrated in the hands of very few people. One way to address this is to look at GDP alongside another measure of economic development, such as the Human Development Index (HDI).

The Bottom Line

In their seminal textbook *Economics*, Paul Samuelson and William Nordhaus neatly sum up the importance of the national accounts and GDP. They liken the ability of GDP to give an overall picture of the state of the economy to that of a satellite in space that can survey the weather across an entire continent.

GDP enables policy-makers and central banks to judge whether the economy is contracting or expanding, whether it needs a boost or restraint, and if a threat such as a recession or inflation looms on the horizon. Like any measure, GDP has its imperfections. In recent decades, governments have created various nuanced modifications in attempts to increase GDP accuracy and specificity. Means of calculating GDP have also evolved continually since its conception to keep up with evolving measurements of industry activity and the generation and consumption of new, emerging forms of intangible assets.

India - GDP

Economic growth hits a record high in Q2

Economic growth surged in the second quarter of the calendar year 2021 (Q1 FY 2021), with GDP expanding 20.1% on an annual basis (Q1 CY 2021: +1.6% year-on-year). Q2's reading marked the strongest increase since current records began and was largely the result of a low base effect. Consequently, output was still down 9.2% compared to the same period in CY 2019.

The upturn reflected improvements in private consumption and fixed investment. Private consumption growth increased to 19.2% year-on-year (Q1 CY 2021: 2.7% yoy). Fixed investment growth picked up to 55.3% in the quarter from 10.9% in Q1 CY 2021. In contrast, government consumption declined 4.8% in Q2 (Q1 2021: +28.3% yoy).

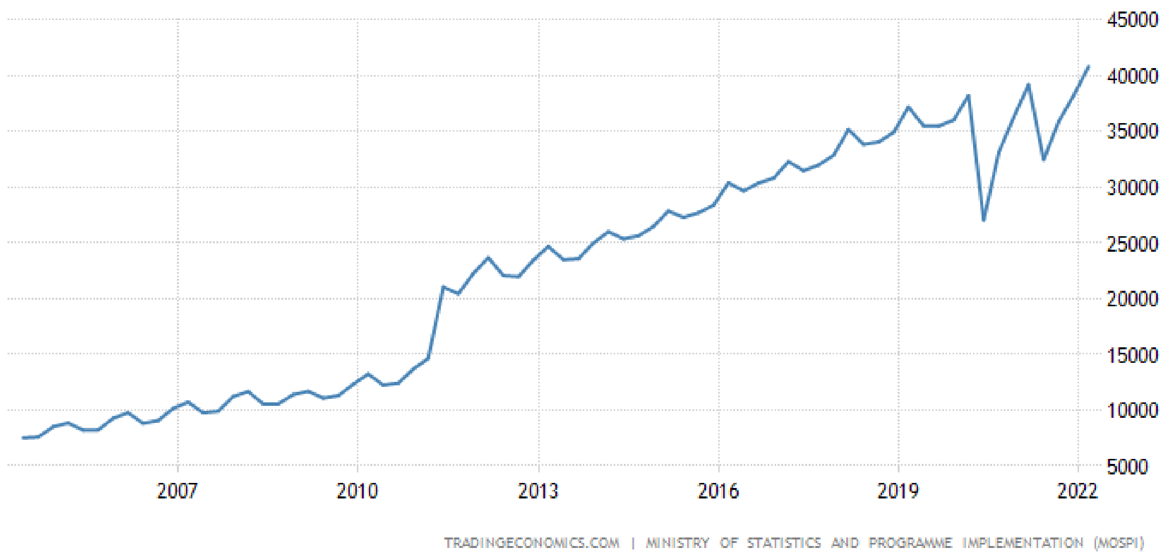
On the external front, exports of goods and services rallied, jumping 39.1% year-on-year in the second quarter (Q1 CY 2021: 8.8% yoy). In addition, imports of goods and services growth accelerated to 60.2% in Q2 (Q1 CY 2021: 12.3% yoy). As a result, the external sector subtracted 3.6 percentage points from growth, which was stronger than the 1.0 percentage-point contribution in the prior quarter.

Looking to H2 CY 2021 (H2 FY 2022), underlying growth momentum is unlikely to be sustained, due to still-elevated uncertainty surrounding the pandemic and as the impact from the low base effect fades. Nevertheless, activity should still show some resilience and grow robustly for the remainder of this calendar year as some easing of restrictions in recent months supports the services sector.

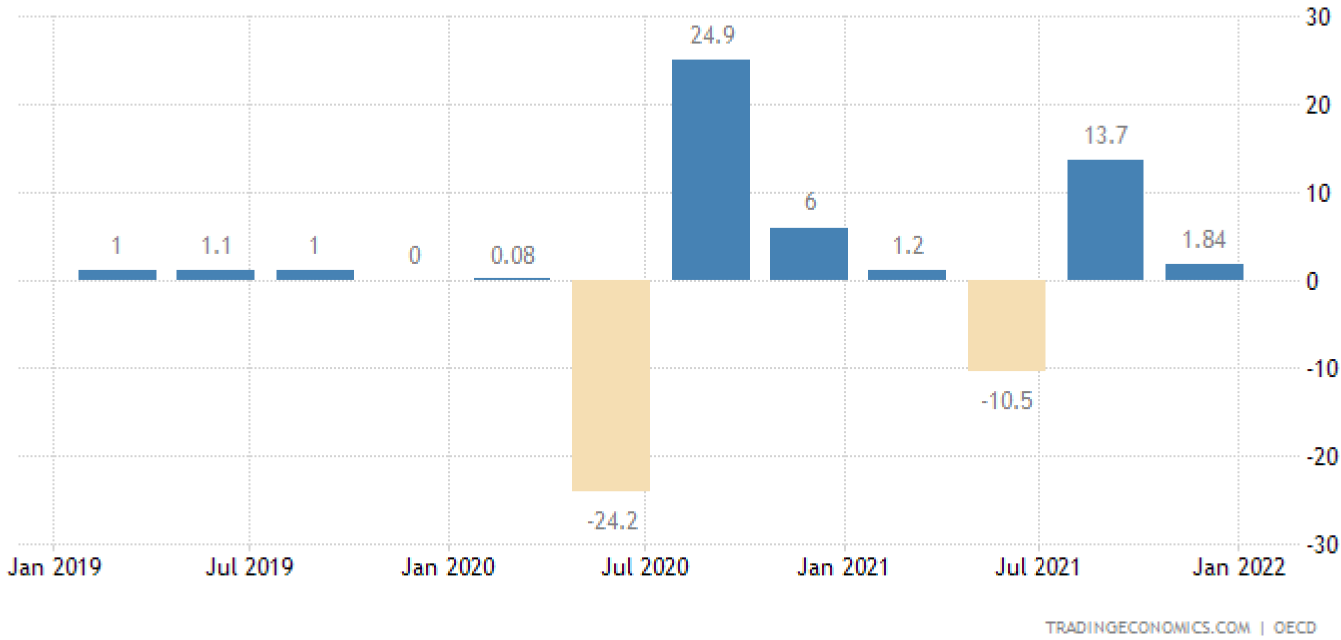
Commenting on India's GDP outlook, Barnabas Gan, an economist at UOB, noted:

"In a nutshell, India's growth prospects will depend largely on how Covid-19 evolves. India's GDP had expanded strongly from its full-year contraction of 7.3% in FY2020/21, and anecdotal evidence from lower Covid-19 infections and higher vaccination rates are credible signals that the economy is gearing towards a more resilient growth pattern. On the back of an accommodative monetary policy expected in the year ahead, coupled with a strong fiscal response as seen from the Union Budget, we keep to our full-year growth outlook of 8.5% in FY2021/22."

GDP AT CONSTANT PRICE



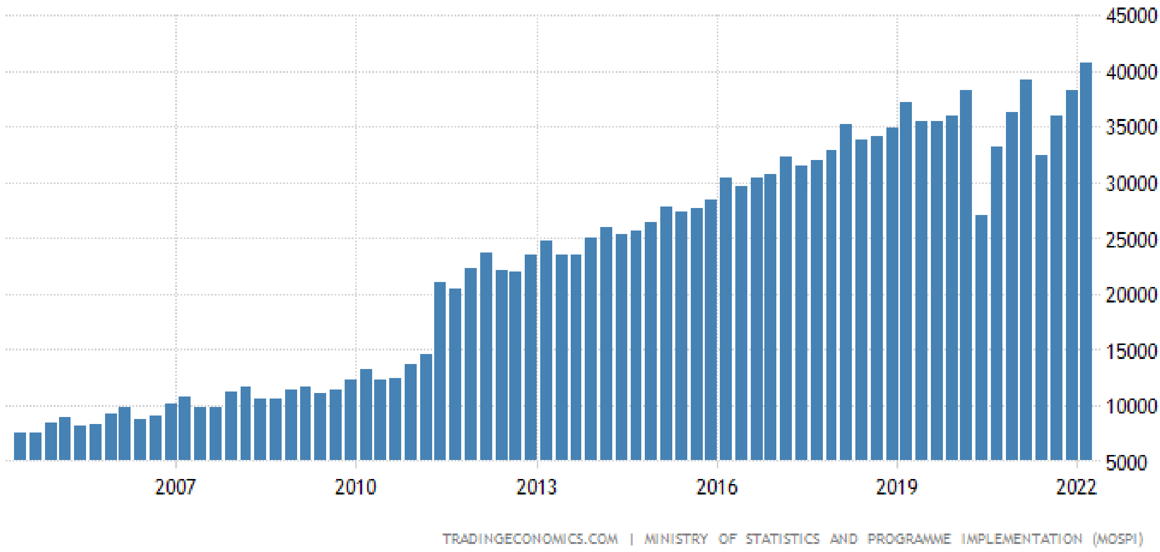
FocusEconomics panelists project GDP to expand 9.0% in FY 2021, which is down 0.2 percentage points from last month's forecast, and increase 7.3% in FY 2022, which is up 0.2 percentage points from the previous month's estimate. The Gross Domestic Product (GDP) in India expanded 1.84 percent in the fourth quarter of 2021 over the previous quarter. source: OECD



Last	Previous	Unit	Reference
1.84	13.70	percent	Dec 2021
4.10	5.40	percent	Mar 2022
2622.98	2870.50	USD Billion	Dec 2020
40780.25	38218.78	INR Billion	Mar 2022
8.70	-6.60	percent	Mar 2022

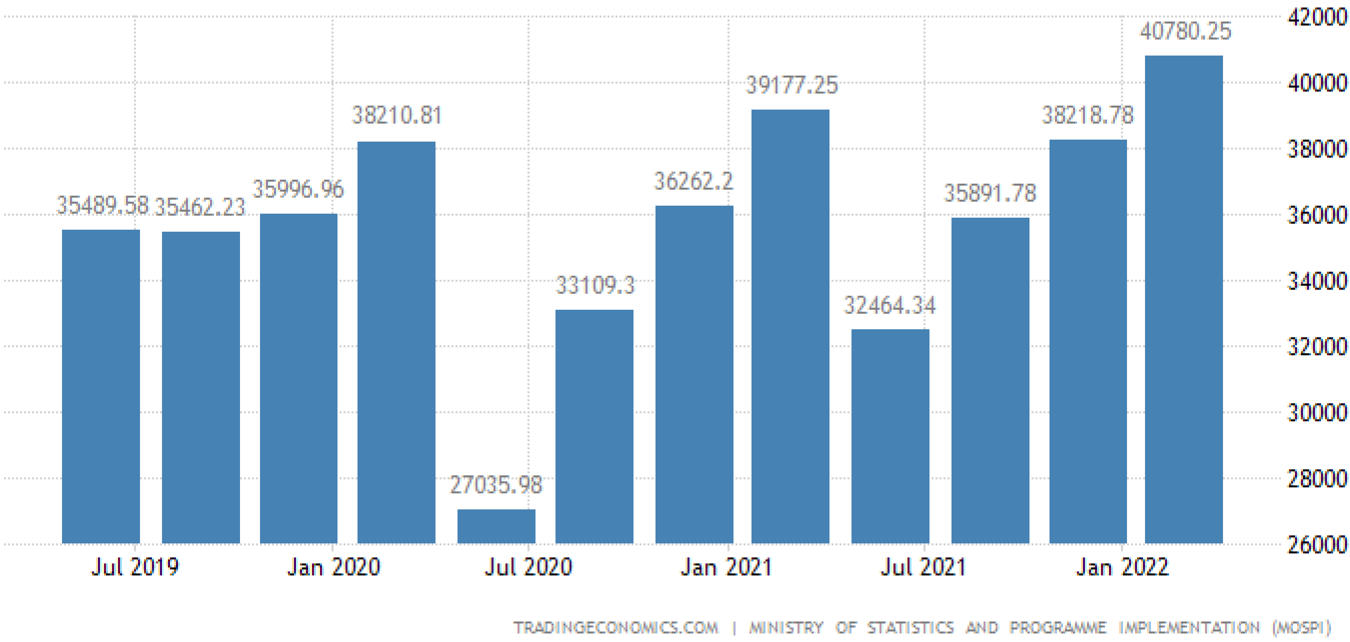
India GDP Growth Rate

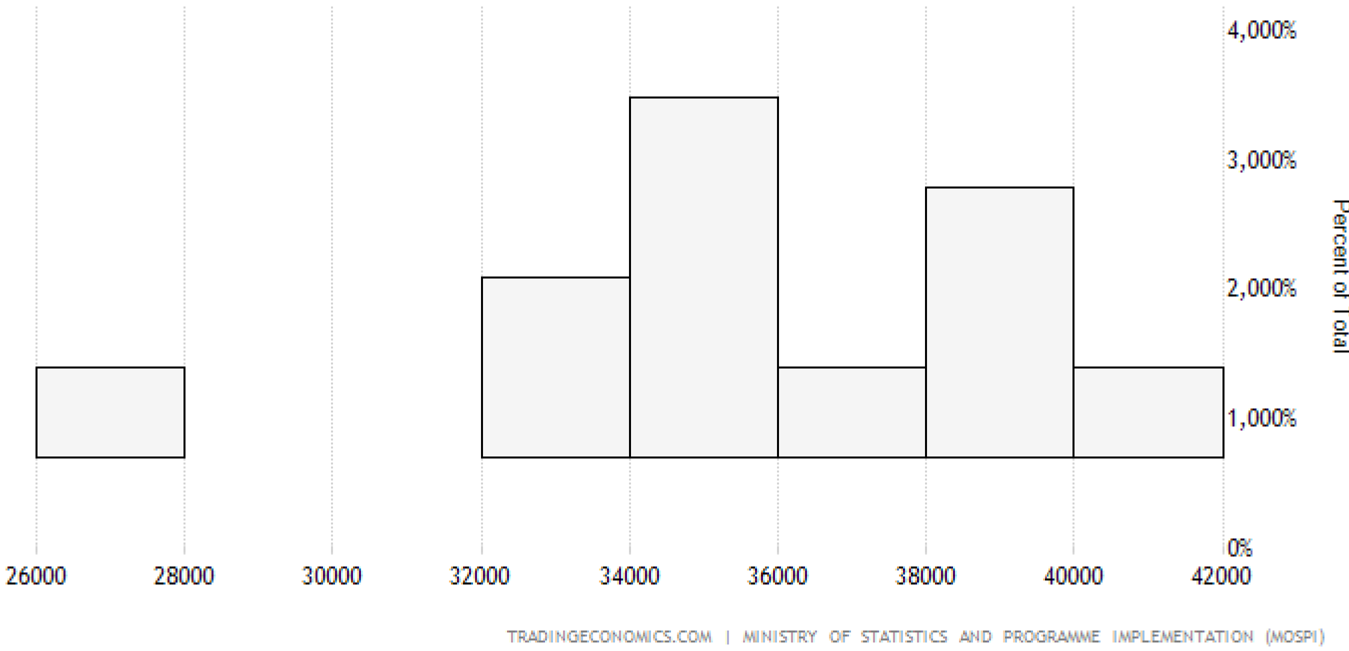
In India, the growth rate in GDP measures the change in the seasonally adjusted value of the goods and services produced by the Indian economy during the quarter. India is the world’s tenth largest economy and the second most populous. The most important and the fastest growing sector of Indian economy are services. Trade, hotels, transport and communication; financing, insurance, real estate and business services and community, social and personal services account for more than 60 percent of GDP. Agriculture, forestry and fishing constitute around 12 percent of the output, but employs more than 50 percent of the labor force. Manufacturing accounts for 15 percent of GDP, construction for another 8 percent and mining, quarrying, electricity, gas and water supply for the remaining 5 percent.



1.84	13.70	24.90	-24.20	1996 - 2021	percent	Quarterly	SA

GDP Constant Prices in India increased to 40780.25 INR Billion in the first quarter of 2022 from 38218.78 INR Billion in the fourth quarter of 2021. source: Ministry of Statistics and Programme Implementation (MOSPI)





Advantages of GDP

- It is easily the best indicator of development in terms of how broad it's reach is and also the expenditures that it covers.

- It is very easy to measure GDP in terms of percentage and it also forms a very reliable source of information when comparing with one's previous years GDP and with other countries.
- As the calculations methods are used all over the world by countries to calculate their economies, so it can be regarded as a universal safe indicator.
- It is a good way for governments to know whether economic policies have been successful or not, and to what extent they have or have not been.
- It is a cheap method in comparison and also the process of collection of data is easy which makes it working smoothly and non-troublesome.

Criticisms of GDP

There are, of course, drawbacks to using GDP as an indicator. In addition to the lack of timeliness, some criticisms of GDP as a measure are:

- **It ignores the value of informal or unrecorded economic activity**
- **It is geographically limited in a globally open economy:**
- **It emphasizes material output without considering overall well-being:**
- **It ignores business-to-business activity**
- **It counts costs and waste as economic benefits**.

Conclusion

The various advantages tell us why GDP is the better method for calculating economic development. Despite these laurels under the method’s belt, it has been criticised for not being able to measure quality of life, standard of living and even political freedom among many others.

This just proves that no method comes with a cent percent guarantee of being accurate but still GDP is a very effective method and in terms of finance and economic growth it might even be the best.

2) Autoregressive Integrated Moving Average (ARIMA) model:-

ARIMA is a type of model known as a Box-Jenkins method. ARIMA is a time series data forecasting statistical model. The ARIMA equation is a regression type equation with lags of the dependent variable and/or lags of forecast errors as independent variables. The ARIMA model's equation is as follows:

$y'(t) = c + \varphi 1 * y'(t-1) + \cdots + \varphi p * y'(t-p) + \theta 1 * \varepsilon(t-1) + \cdots + \theta q * \varepsilon(t-q) + \varepsilon t$

An ARIMA model is characterized by 3 terms: p, d, q

where,

p is the order of the AR term

q is the order of the MA term

d is the number of differencing required to make the time series stationary.

The ARIMA model's popularity stems from its ability to represent a wide range of time series with ease, as well as the accompanying Box -Jenkins approach for an optimal model construction process. However, the pre-assumed linear form of the associated time series is a serious restriction of these models.In many real-life circumstances, this is insufficient. Various methods have been proposed to address this disadvantage.

The ARMA(Auto Regressive Moving Average) models can only be used for stationary time series data. However in practice many time series such as those related to socio-economic business show non-stationary behaviour. Time series, which contain trend and seasonal patterns, are also non-stationary in nature. Thus from application view point , ARMA models are inadequate to properly describe non-stationary time series, which are frequently encountered in practice. For this reason the ARIMA model is proposed, which is a generalization of an ARMA model to include the case of non-stationarity as well. In ARIMA models a non-stationary time series is made stationary by applying finite differencing of the data points. An ARIMA(p,0,0) is nothing but the AR(p) model and ARIMA(0,0,q) is the MA(q) model. ARIMA(0,1,0), is a special one and known as the Random Walk model . ARIMA models are applied in the cases where the data shows evidence of non-stationarity. In time series analysis, non-stationary data are always transformed into stationary data.

The common causes of non-stationary in time series data are the trend and the seasonal components. The way to transformed non-stationary data to stationary is to apply the differencing step. It is possible to apply one or more times of differencing steps to eliminate the trend component in the data. Similarly, to remove the seasonal components in data, seasonal differencing could be applied.

According to the name, we can split the model into smaller components as follow:

AR: an Auto Regressive model which represents a type of random process. The output of the model is linearly dependent on its own previous value i.e. some number of lagged data points or the number of past observations .

MA: a Moving Average model which output is dependent linearly on the current and various past observations of a stochastic term.

I: integrated here means the differencing step to generate stationary time series data, i.e. removing the seasonal and trend components

For seasonal time series forecasting, Box and Jenkins had proposed a quite successful variation of ARIMA model, viz. the Seasonal ARIMA (SARIMA)

2.1) Stationary Time Series

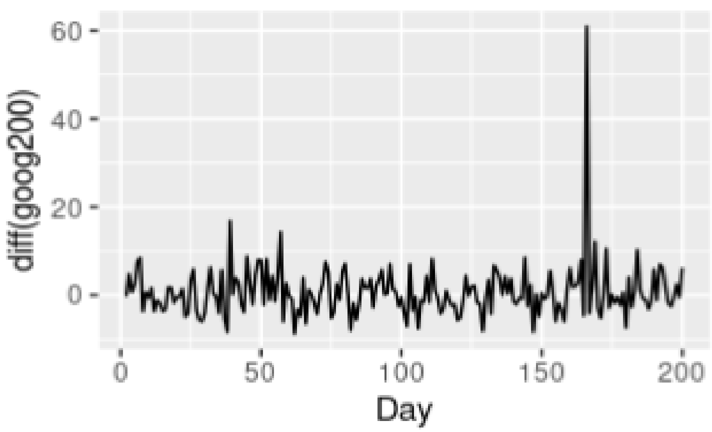
A key role in time series analysis is played by processes whose properties, or some of them, do not vary over time. Such a property is illustrated in the following important concept, stationarity. We then introduce the most commonly used stationary linear time series models–the autoregressive integrated moving average (ARIMA) models. These models have assumed great importance in modelling real-world processes.

A stationary time series is one whose properties do not depend on the time at which the series is observed.

Time series containing trends or seasonality, on the other hand, are not stationary since the trend and seasonality will alter the value of the time series at different times.

Some cases can be confusing — a time series with cyclic behaviour (but with no trend or seasonality) is stationary. This is because the cycles are not of a fixed length, so before we observe the series we cannot be sure where the peaks and troughs of the cycles will be.

In general, a stationary time series will have no predictable patterns in the long-term. Time plots will show the series to be roughly horizontal (although some cyclic behaviour is possible), with constant variance.



2.2) DIFFERENCING:-

One way to make a non-stationary time series stationary — compute the differences between consecutive observations. This is known as differencing.

Logarithms and other transformations can aid in the stabilization of a time series' volatility. By removing fluctuations in the level of a time series and so eliminating (or lowering) trend and seasonality, differencing can assist stabilize the mean of a time series.

Second-order differencing

Occasionally the differenced data will not appear to be stationary and it may be necessary to difference the data a second time to obtain a stationary series

Looking at the time plot of the data, the ACF plot is also useful for identifying non-stationary time series. For a stationary time series, the ACF will drop to zero relatively quickly, while the ACF of non-stationary data decreases slowly.

An autoregressive integrated moving average – ARIMA model is a generalization of a simple autoregressive moving average – ARMA model. Both of these models are used to forecast or predict future points in the time-series data. ARIMA is a form of regression analysis that indicates the strength of a dependent variable relative to other changing variables.

The final objective of the model is to predict future time series movement by examining the differences between values in the series instead of through actual values. ARIMA models are applied in the cases where the data shows evidence of non-stationarity. In time series analysis, non-stationary data are always transformed into stationary data.

According to the name, we can split the model into smaller components as follow:

AR Model:-

An Auto Regressive model which represents a type of random process. The output of the model is linearly dependent on its own previous value i.e. some number of lagged data points or the number of past observations . In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. The term autoregression indicates that it is a regression of the variable against itself.

$$\hat{y}_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t$$

Here, p = the number of lagged observations in the model, ε is white noise at time t, c is a constant and ϕs are parameters.

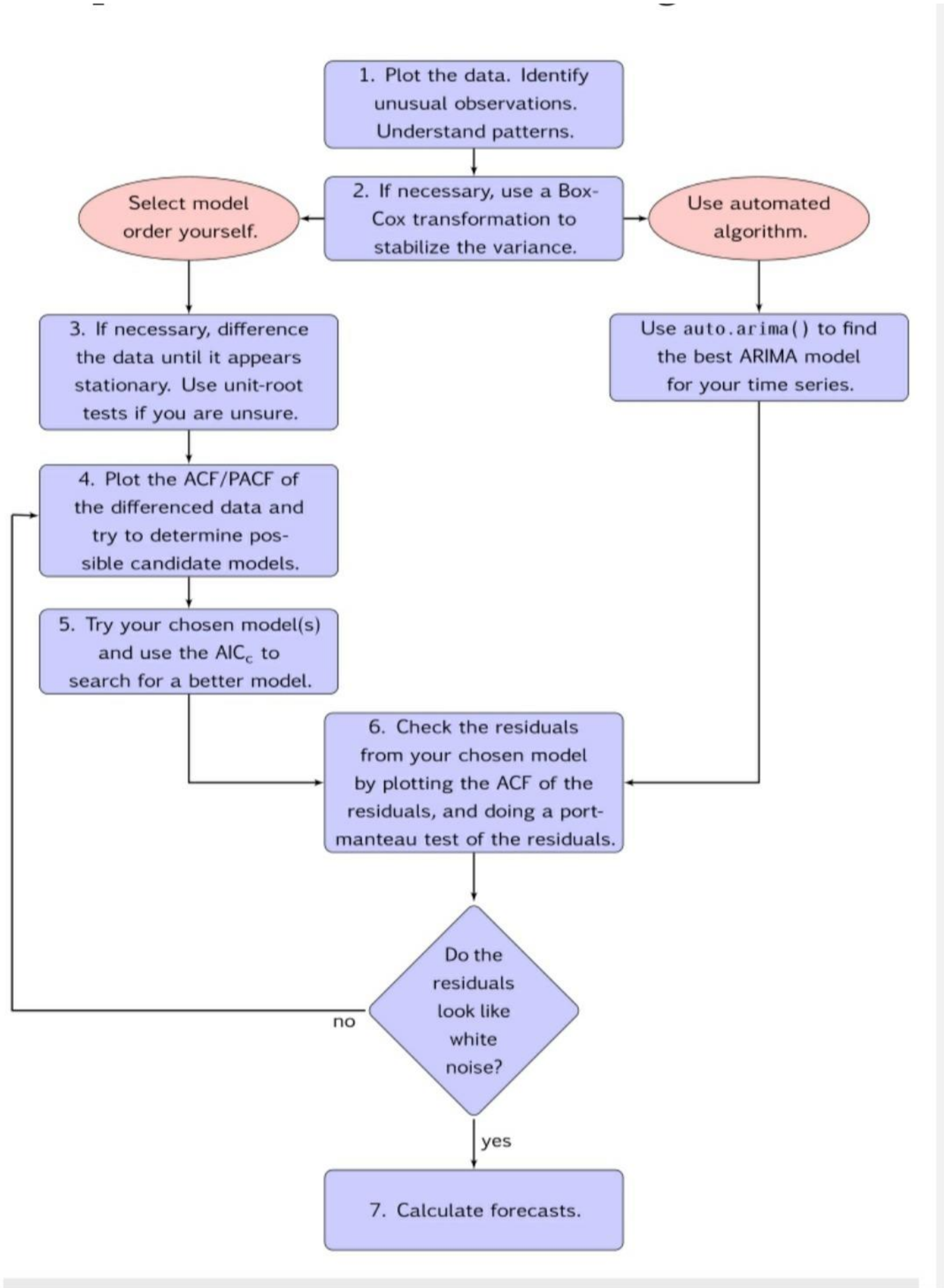
MA Model:-

A Moving Average model which output is dependent linearly on the current and various past observations of a stochastic term. In a regression-like model, a moving average model uses prior prediction mistakes rather than historical values of the forecast variable.

$$yt=c+\varepsilon t+\theta 1\varepsilon t-1+\theta 2\varepsilon t-2+\_\_+\theta q\varepsilon t-q$$

Here, ε is white noise at time t, c is a constant, and θs are parameters

Process For Forecasting In Arima Model:-



ACF: The autocorrelation coefficient function, define how the data points in a time series are related to the preceding data points.

The Autocorrelation Function (ACF) is used to determine the number of MA(q) terms in the model. It determines the correlation between the observations at the current point in time and all previous points in time.

ACF plot shows the autocorrelations which measure the relationship between

$Y_t$and y_{t-k} for different values of k. Now if $y_t$and y_{t-1} and

Y_{t-1} and y_{t-2} must also be correlated. However, then

y_t and y_{t-2} might be correlated, simply because they are both connected to y_{t-1} , rather than because of any new information contained in

y_{t-2} that could be used in forecasting $y_{t.}$

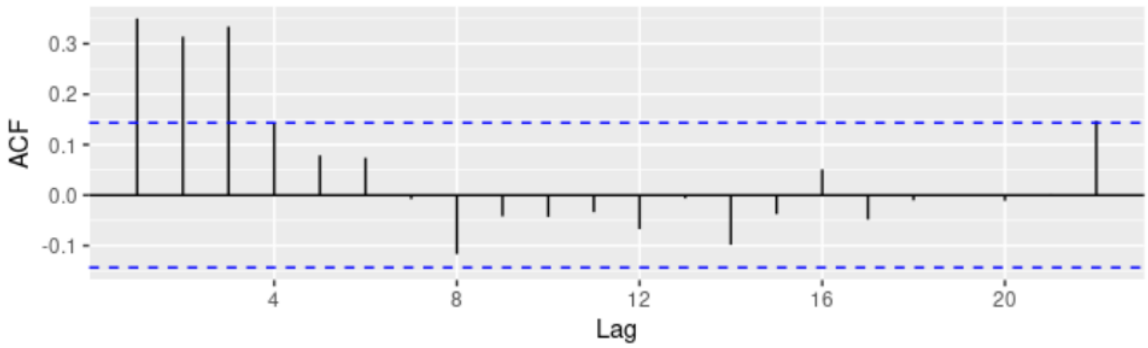


Figure 8.9: ACF of quarterly percentage change in US consumption.

PACF: The partial autocorrelation coefficient function, like the autocorrelation function, conveys vital information regarding the dependence structure of a stationary process.

The Partial Autocorrelation Function (PACF) results determine the order of the model or the values for the MA portion of the model.

Now let’s explain how to select a model based on ACF and PACF plots.

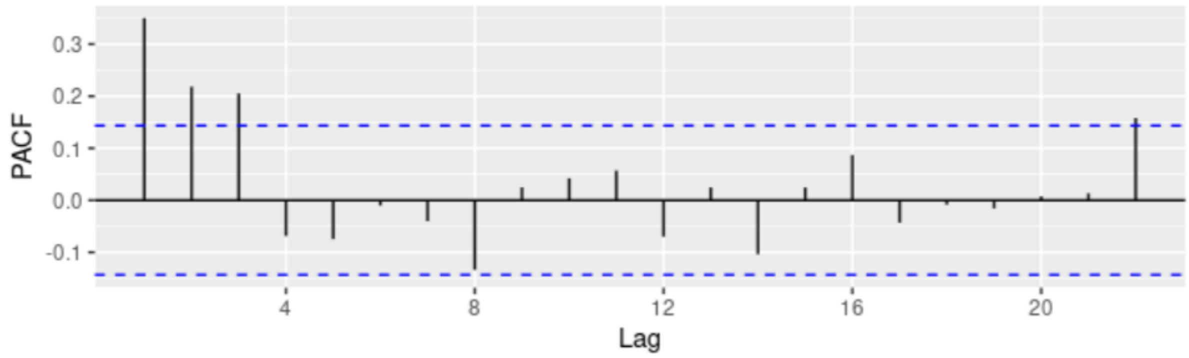


Figure 8.10: PACF of quarterly percentage change in US consumption.

Given observations of a time series, one approach to the fitting of a model to the data is to match the sample ACF and sample PACF of the data with the ACF and PACF of the model, respectively. We can use the sample ACF plot to see if the specific time series data is stationary or not. Here is an example of how the sample ACF plot will look like if the time series is not stationary.

For stationary time-series data, the sample ACF will die down very quickly or even cuts off. Otherwise, the time series is not stationary. Cut off here means the ACF value is less than the indicated confidence interval or inside the blue dotted line.

If we confirm the data is stationary, we can decide which parameters q and p should be used for the model based on at which lag the sample ACF and PACF value cut off. In this case, parameter p will be decided by sample PACF cut off time and parameter q will be decided by sample ACF cut off time. If we know the time series is not stationary, will try to perform the higher-order differencing to ensure the data is stationary.

Lag

The lag time is the time between the twotime series you are correlating. If you have time series data at t=0,1,…,n, then taking the autocorrelation of data sets (0,1),(1,2)…(n−1,n) apart would have a lag time of 1. If you took the autocorrelation of data sets (0,2),(1,3),(n−2,n) that would have lag time 2 etc. And autocorrelation is a measure of how much the data sets at one point in time influences data sets at a later point in time.

Maximum Likelihood Estimation:-

When R estimates the ARIMA model, it uses maximum likelihood estimation (MLE). This technique finds the values of the parameters which maximize the probability of obtaining the data that we have observed.

Auto.arima

Although ARIMA is a very powerful model for forecasting time series data, the data preparation and parameter tuning processes take a long time. Before you can use ARIMA, you must first make the series stationary and calculate the values of p and q using the plots we discussed earlier. Auto ARIMA simplifies this task for us.

We have been going through the process of manually fitting different models and deciding which one is best. So, we are going to automate the process. Basically, it takes the data and fits many models in a different order before comparing the characteristics. However, the processing time increases substantially, when we try to fit complex models. How the model figures out the best combination of these parameters? Auto ARIMA takes into the AIC and BIC values generated (as you can see in the code) to determine the best combination of parameters. AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values are estimators to compare models. The lower these values, the better is the model.

Pros of auto.arima:

- It saves an enormous amount of time
- Eliminate the need to understand the statistics and theory behind the model selection.

This method will also reduce the risk of human error and the possible mistakes caused by an incorrect interpretation of the results.

Cons of auto.arima:

- Blindly putting our faith into one criterion
- Never really see how well the other models perform
- Topic expertise

Box.Test

The Ljung Box test is a modification of the Box Pierce Test. The difference is in how the test statistic is calculated.

The Box-Ljung test is a diagnostic tool for determining whether a time series model is unfit. The test is applied to the residuals of a time series after fitting an ARMA((p,q)) model to the data. The test looks at residual (m) autocorrelations. If the autocorrelations are very small, we conclude that the model does not have a significant lack of fit. The Ljung (pronounced Young) Box test (sometimes called the modified Box-Pierce, or just the Box test) is a way to test for the absence of serial autocorrelation, up to a specified lag k.

Alternative ways to test for autocorrelation:-

The Durbin-Watson test is a popular way to test for autocorrelation. However, it can’t be used if you have lagged dependent variables; If you have those, used the Breusch-Pagan-Godfrey test instead.

L-jung Box Test Hypothesis

The null hypothesis of the Box Ljung Test, H_0, is that our model does not show lack of fit (or in simple terms—the model is just fine). The alternate hypothesis, H_a, is just that the model does show a lack of fit.

Residual

A line of best fit is obtained when performing simple linear regression (or any other type of regression analysis). The data points do not always fall exactly on this regression equation line; they are dispersed. The vertical distance between a data point and the regression line is defined as a residual. There is one residual for each data point. They are:

Positive if they are above the regression line,

Negative if they are below the regression line,

Zero if the regression line actually passes through the point,

Because residuals are the difference between any data point and the regression line, they are also referred to as "errors" in some contexts. In this context, error does not imply that there is something wrong with the analysis;

it simply indicates that there is some unexplained difference. In other words, the residual is the error that the regression line does not explain.

Residual = Observed value – predicted value

$e = y - \hat{y}$

FORECASTING

Forecasting is the process of using models fit on historical data to predict future observations. A key distinction in forecasting is that the future is completely unknown and can only be estimated based on what has already occurred. Time series models have numerous and diverse applications, ranging from sales forecasting to weather forecasting. Time Series Forecasting attempts to forecast a future value or classification at a specific point in time.

Time series forecasts inform all kinds of business decisions. Some examples:

- Forecasting infection rates to improve disease control and outbreak prevention programmes.
- Predicting customer ratings for forecasting product sales.
- Forecasting power demand to determine whether another power generation plant should be built in the next five years.

Potential pros of using ARIMA models

- Only requires the prior data of a time series to generalize the forecast.
- Performs well on short term forecasts.
- Models non-stationary time series.

Potential cons of using ARIMA models

- Difficult to predict turning points.
- There is quite a bit of subjectivity involved in determining (p,d,q) order of the model.
- Computationally expensive.
- Poorer performance for long term forecasts.
- Cannot be used for seasonal time series.
- Less explainable than exponential smoothing.

3. Methodology

The following steps is the description of the steps involved in the sequence. At the end, the evaluation is performed and the final results are obtained.

a) **Data collection**. We have collected the GDP growth data for India from 1951-2020 from a commaseparated (CSV) file available for download on the official website of Open Government Data (OGD) Platform India, and uploaded that in Excel 2013.Then data was cleaned , extracted and converted into Time series data to be used for analysis and forecasting in R studio software.

b)**Analysis of data-** Data was analyzed for missing values and plotted to visualize the GDP growth of India from 1951-2020.

c) **Time Series forecasting**. Time series forecasting with ARIMA Model is used for forecasting the trend of GDP growth in the near future, i.e. for next ten years from 2020-2030.

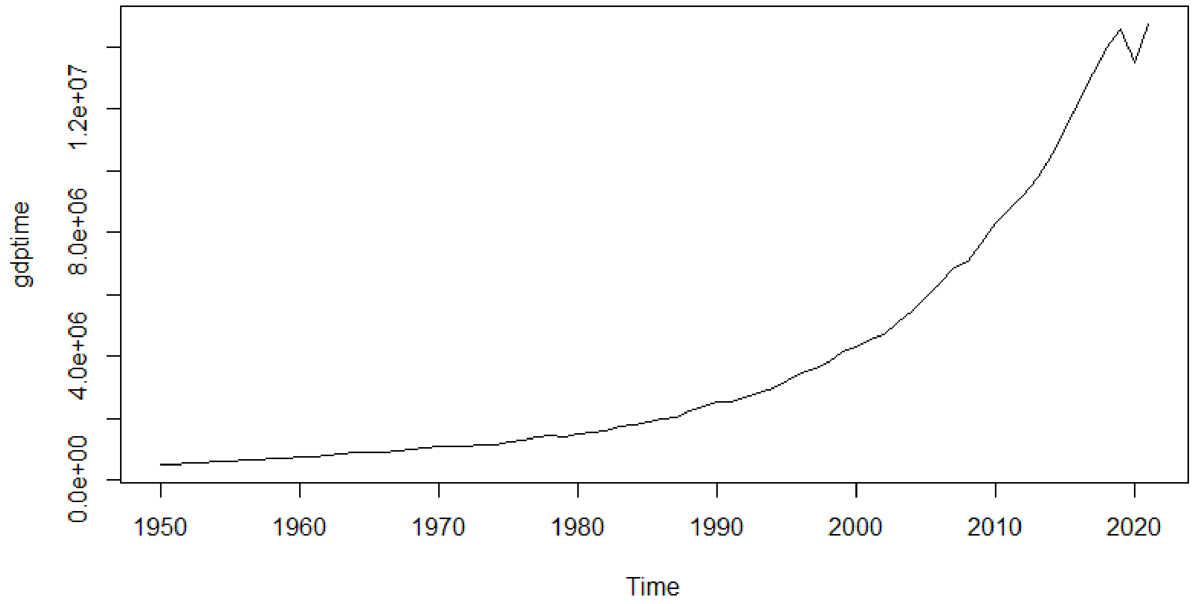
d)**Validation of results.** Models are validated on the basis of comparison of actual test data with the forecasted data by dividing the dataset into a training set and testing set. The model prediction is then tested on test data for ten years with actual data to check the predictive capability of the model.

4) Statistical Analysis

4.1) #CODING FOR ANNUALLY GDP DATA AT CONSTANT PRICE

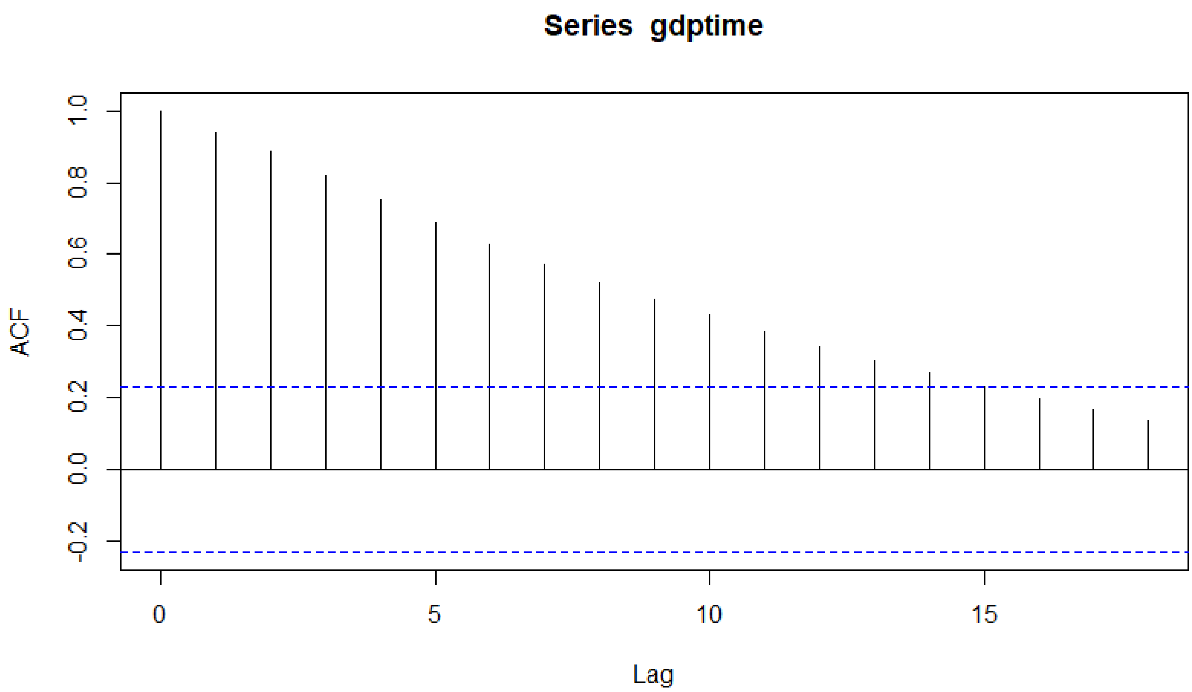
```
#Load Libraries
  library("tseries")
library(forecast)
 #Imoprt Data
    library(readxl)
  gdp <- read_excel("OneNote Notebooks/New folder/gdp.xlsx")
  View(gdp)
  attach(gdp)
  summary(gdp)
year          GDP
Min.   :1950   Min.   :  496848
1st Qu.:1968   1st Qu.:  964261
Median :1986   Median : 1928220
Mean   :1986   Mean   : 3829830
3rd Qu.:2003   3rd Qu.: 5178632
Max.   :2021   Max.   :14753535
  #convert To Time Series
  gdptime<-ts(gdp$GDP,start = min(gdp$year),end = max(gdp$year),frequency =
1)
```

FIGURE 1-

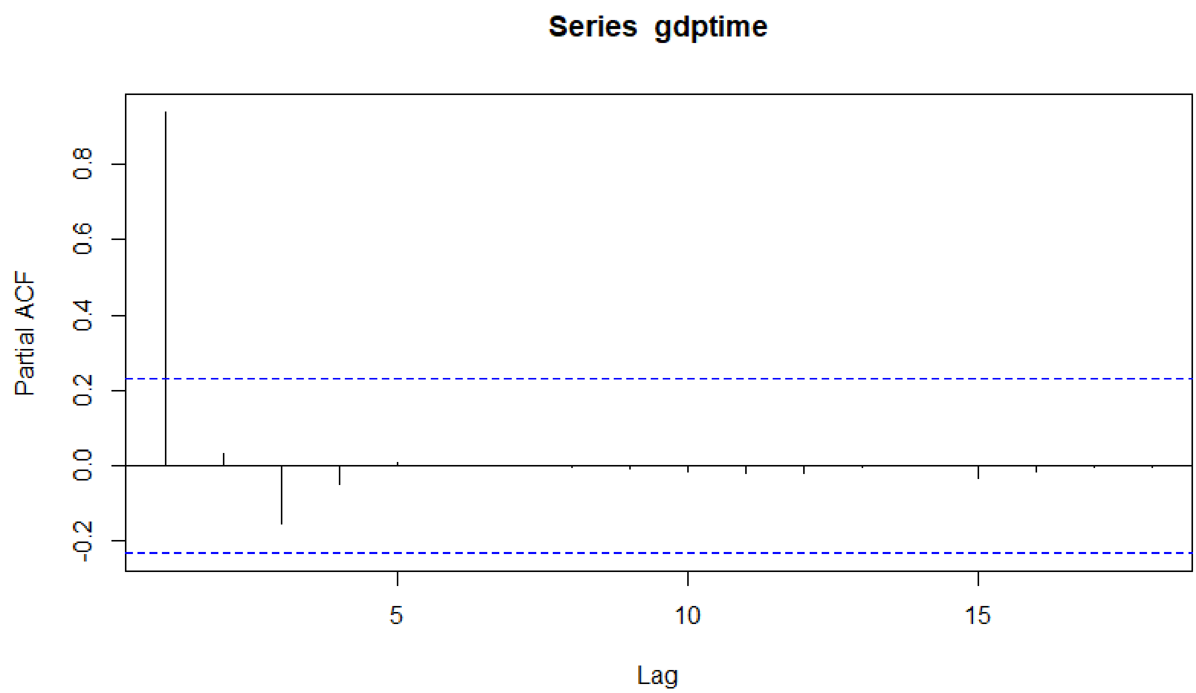


#Check StationarityOfData

acf(gdptime)



pacf(gdptime)



adf.test(gdptime)

Augmented Dickey-Fuller Test

data: gdptime

Dickey-Fuller = 2.5501, Lag order = 4, p-value = 0.99

alternative hypothesis: stationary

In adf.test(gdptime) : p-value greater than printed p-value

#Make Stationary

gdp1=diff(log(gdptime))

adf.test(gdp1)

Augmented Dickey-Fuller Test

data: gdp1

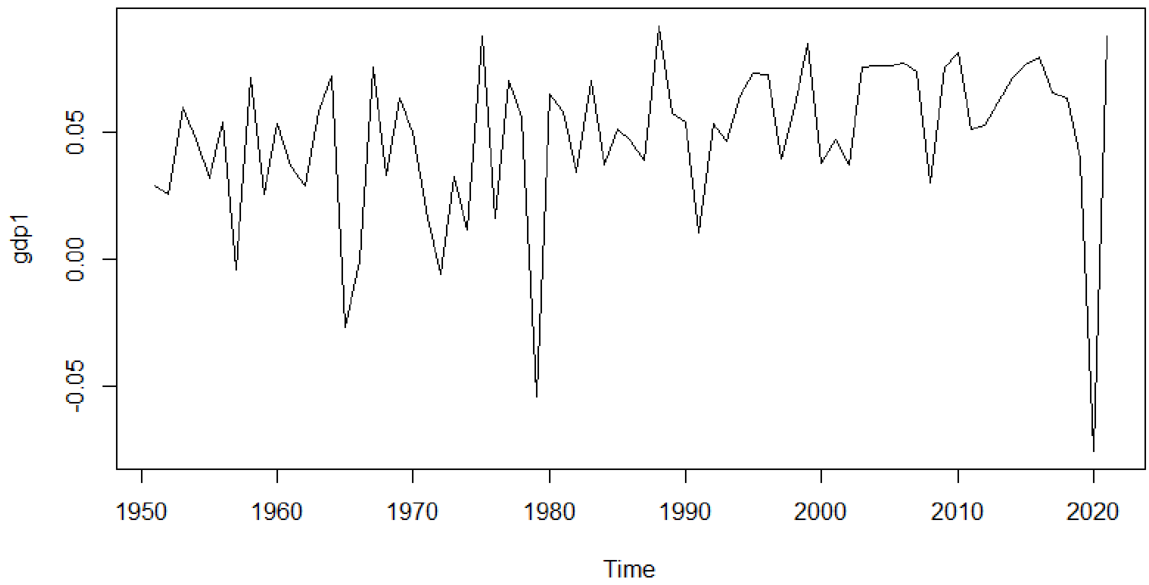
Dickey-Fuller = -4.2826, Lag order = 4, p-value = 0.01

alternative hypothesis: stationary

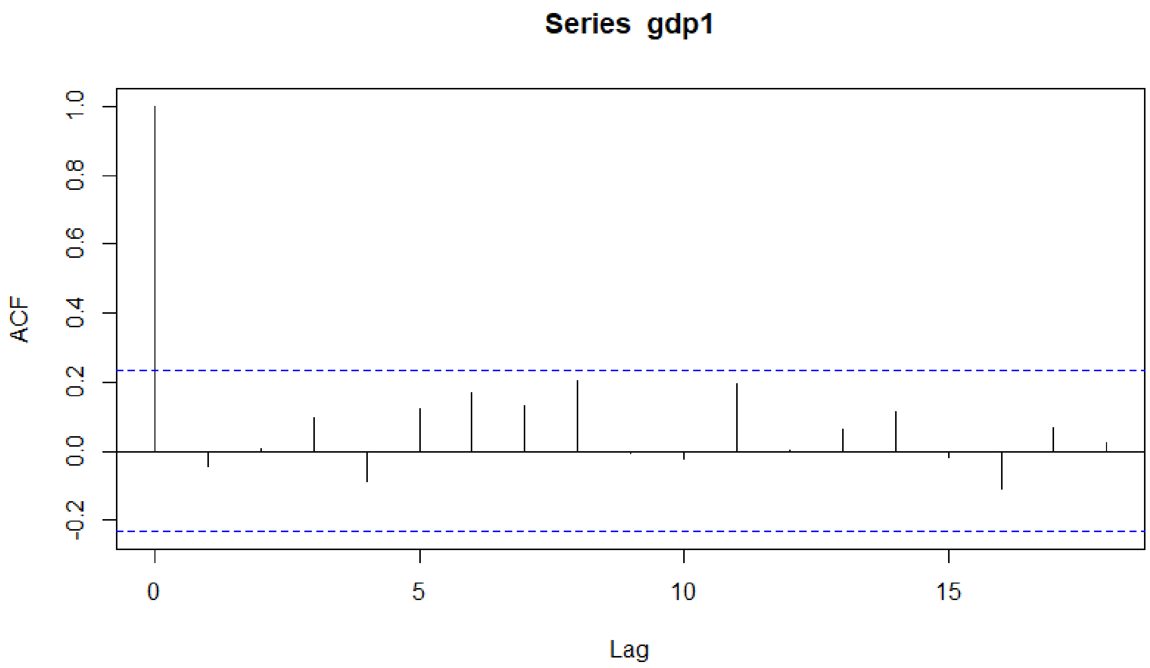
In adf.test(gdp1) : p-value smaller than printed p-value

plot(gdp1)

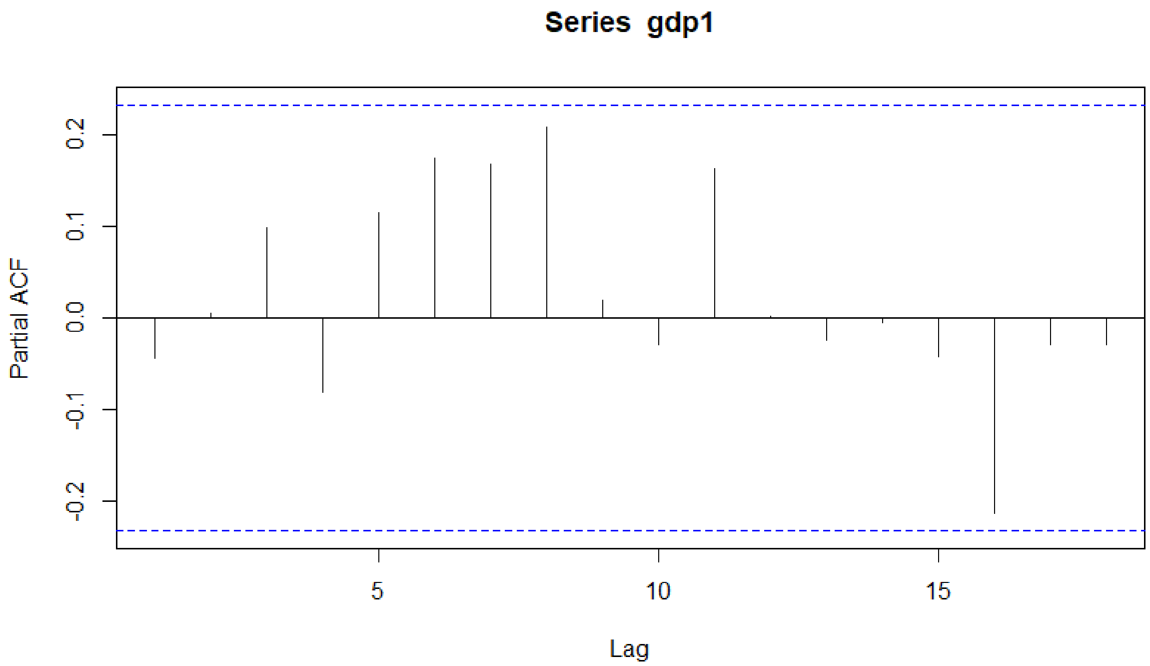
FIGURE -2



acf(gdp1)



pacf(gdp1)



#Fit the Model

```
gdpmodel=auto.arima(gdptime,ic="aic",trace = TRUE)
```

ARIMA(2,2,2) : Inf

ARIMA(0,2,0) : 1987.875

ARIMA(1,2,0) : Inf

ARIMA(0,2,1) : 1948.506

ARIMA(1,2,1) : 1948.18

ARIMA(2,2,1) : 1950.179

ARIMA(1,2,2) : 1950.18

ARIMA(0,2,2) : 1948.206

ARIMA(2,2,0) : 1950.95

Best model: ARIMA(1,2,1)

gdpmodel

Series: gdptime

ARIMA(1,2,1)

Coefficients:

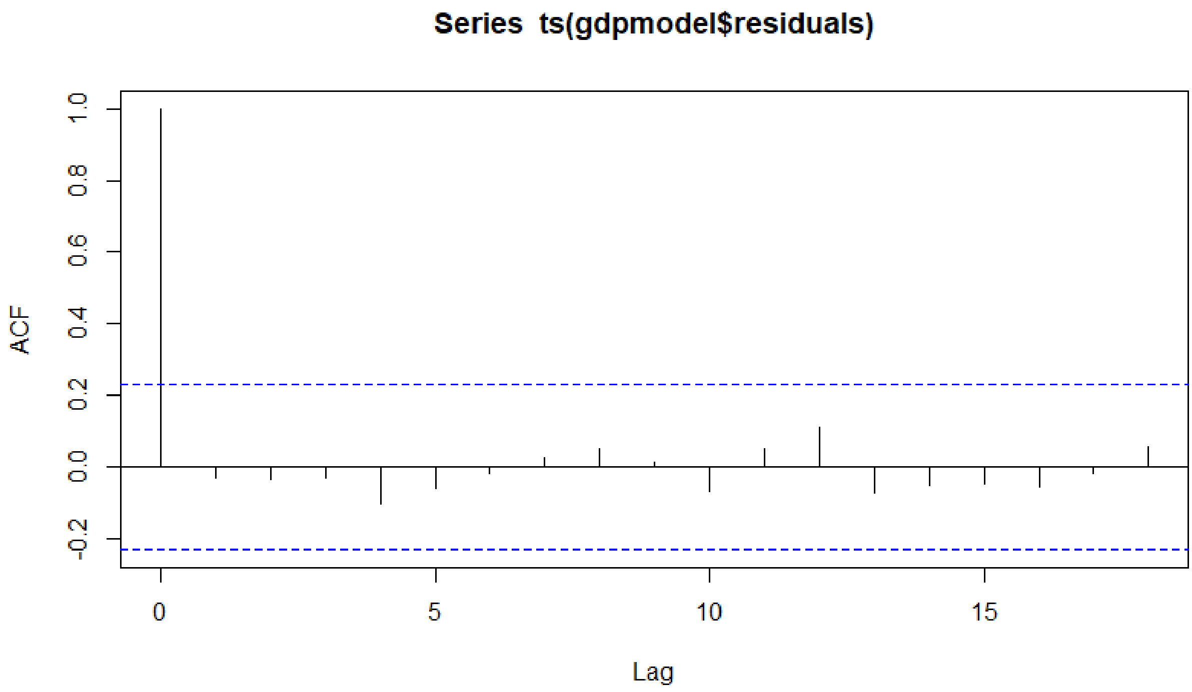
	ar1	ma1
	-0.2399	-0.7690
s.e.	0.1626	0.0955

sigma^2 = 6.634e+10: log likelihood = -971.09

AIC=1948.18   AICc=1948.54   BIC=1954.93

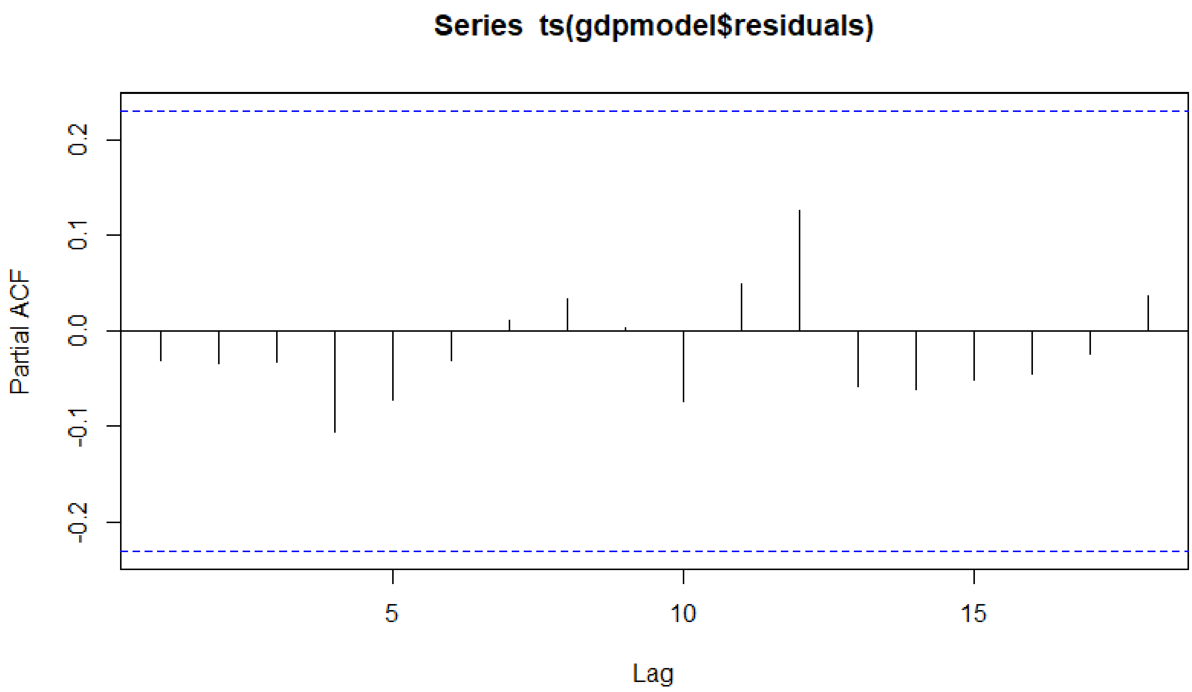
#residuals Diagnosis

acf(ts(gdpmodel$residuals))



The blue line above shows significantly different values than zero.

pacf(ts(gdpmodel$residuals))



#Forecast Future Values

Gdpforecast=forecast(gdpmodel,level = c(95),h=10)

gdpforecast

	Point Forecast	Lo 95	Hi 95
2022	15046839	14542019	15551660
2023	15567475	14856729	16278222
2024	16033568	15091707	16975429
2025	16512747	15335422	17690071
2026	16988786	15564362	18413210
2027	17465578	15783033	19148123
2028	17942190	15990169	19894210
2029	18418845	16186184	20651506
2030	18895489	16371234	21419745
2031	19372136	16545594	22198679

attach(gdpforecast)

summary(gdpforecast)

Forecast method: ARIMA(1,2,1)

Model Information:

Series: gdptime

ARIMA(1,2,1)

Coefficients:

ar1 ma1

-0.2399 -0.7690

s.e. 0.1626 0.0955

sigma^2 = 6.634e+10: log likelihood = -971.09

AIC=1948.18 AICc=1948.54 BIC=1954.93

Error measures:

ME RMSE MAE MPE MAPE MASE ACF1

Training set 34175.4 250309.5 106046.7 0.9543078 2.44862 0.4539256 -0.0301683

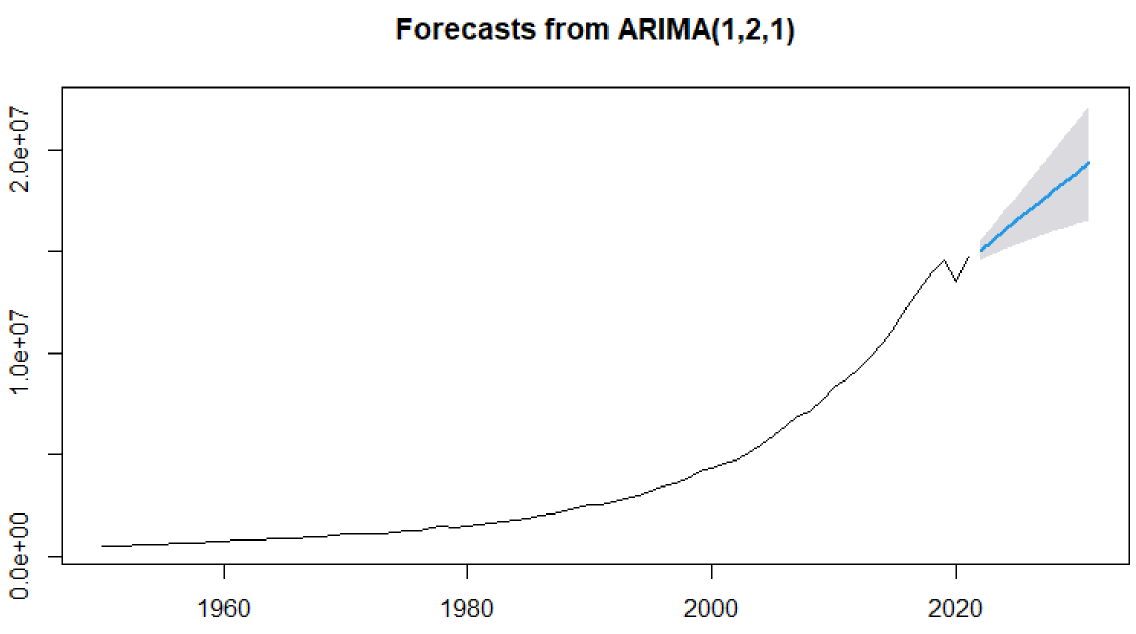
Forecasts:

Point Forecast Lo 95 Hi 95

2022	15046839	14542019	15551660
2023	15567475	14856729	16278222
2024	16033568	15091707	16975429
2025	16512747	15335422	17690071
2026	16988786	15564362	18413210
2027	17465578	15783033	19148123
2028	17942190	15990169	19894210
2029	18418845	16186184	20651506
2030	18895489	16371234	21419745
2031	19372136	16545594	22198679

plot(gdpforecast)

FIGURE -3



#check Autocorrelation,

Box.test(gdpforecast$resid,lag=5,type="Ljung-Box")

Box-Ljung test

data: gdpforecast$resid

X-squared = 1.3393, df = 5, p-value = 0.9308

#CODE FOR QUARTERLY GDP DATA AT CONSTANT PRICE

#Import data

library(readxl)

gdpq<- read_excel("OneNote Notebooks/New folder/gdpq.xlsx")

View(gdpq)

summary(gdpq)

Date GDP

Min.:2005 Min. :12815900

1st Qu.:2009 1st Qu.:18126000

Median :2013 Median :24162700

Mean :2013 Mean :25019768

3rd Qu.:2017 3rd Qu.:32048325

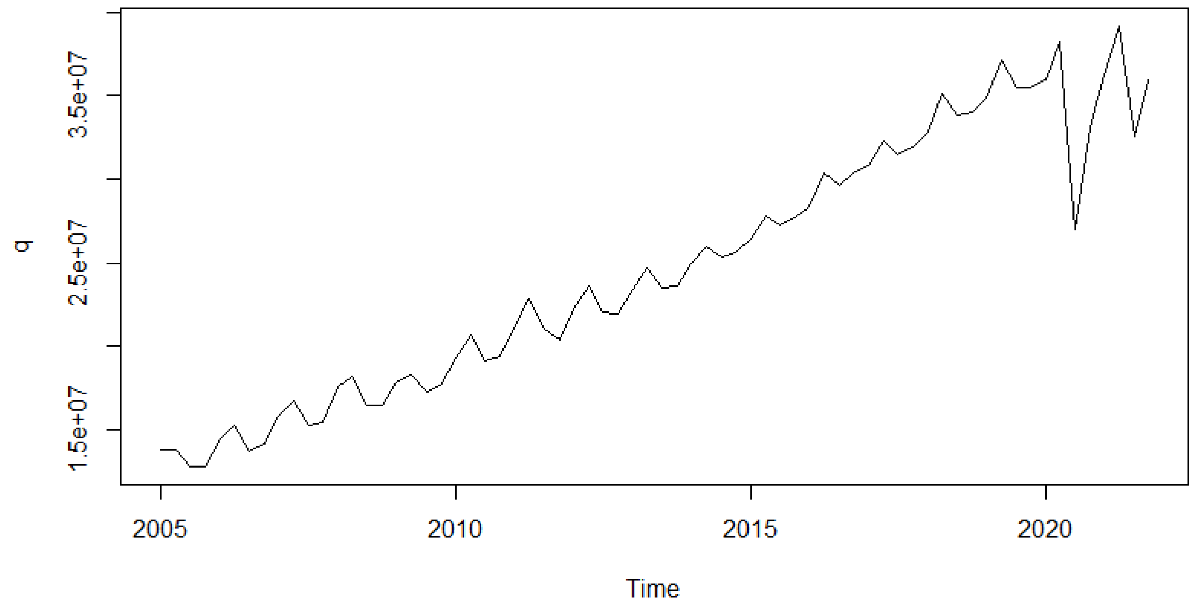
Max. :2021 Max. :39177300

data<-
c(13824500,13824500,12815900,12860700,14497100,15256400,13765000,14177500,15871100,16751500,15253000,15473400,17545100,18229900,16486000,16515600,17814300,18272900,17309300,17664900,19277100,20696300,19091100,19374700,21337900,22833400,21028600,20428700,22251400,23654600,22052200,21959500,23447700,24670800,23474000,23570800,24980000,

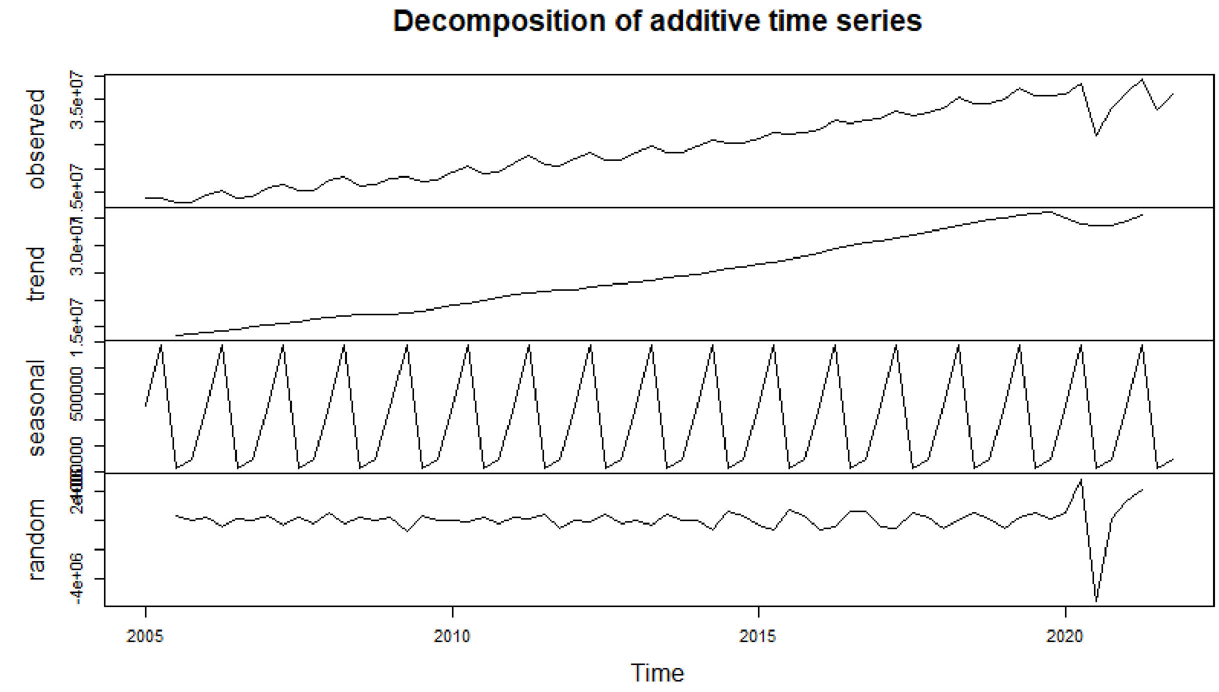
```
25989000,25357500,25622400,26459500,27837300,27282800,27680900,28363900,30367400,29650900,30357600,30796200,32277300,31463400,31972000,32849900,35160600,33823000,34036400,3489760n0,37172100,35489600,35462200,35997000,38210800,27036000,33109300,36262200,39177300,32531100,35916500)

ts(data,frequency = 4,start=c(2005,1))

Qtr1     Qtr2     Qtr3     Qtr4

2005 13824500 13824500 12815900 12860700

2006 14497100 15256400 13765000 14177500

2007 15871100 16751500 15253000 15473400

2008 17545100 18229900 16486000 16515600

2009 17814300 18272900 17309300 17664900

2010 19277100 20696300 19091100 19374700

2011 21337900 22833400 21028600 20428700

2012 22251400 23654600 22052200 21959500

2013 23447700 24670800 23474000 23570800

2014 24980000 25989000 25357500 25622400

2015 26459500 27837300 27282800 27680900

2016 28363900 30367400 29650900 30357600

2017 30796200 32277300 31463400 31972000

2018 32849900 35160600 33823000 34036400

2019 34897600 37172100 35489600 35462200

2020 35997000 38210800 27036000 33109300

2021 36262200 39177300 32531100 35916500

  q<-ts(data,frequency = 4,start=c(2005,1))
```

```
plot(q)
```

FIGURE -4




```
#Decompose Into Time Series Components
timeseriescomponent<-decompose(q)
plot(timeseriescomponent)
```



#Determine Stationarity of Data

adf.test(q)

Augmented Dickey-Fuller Test

data: q

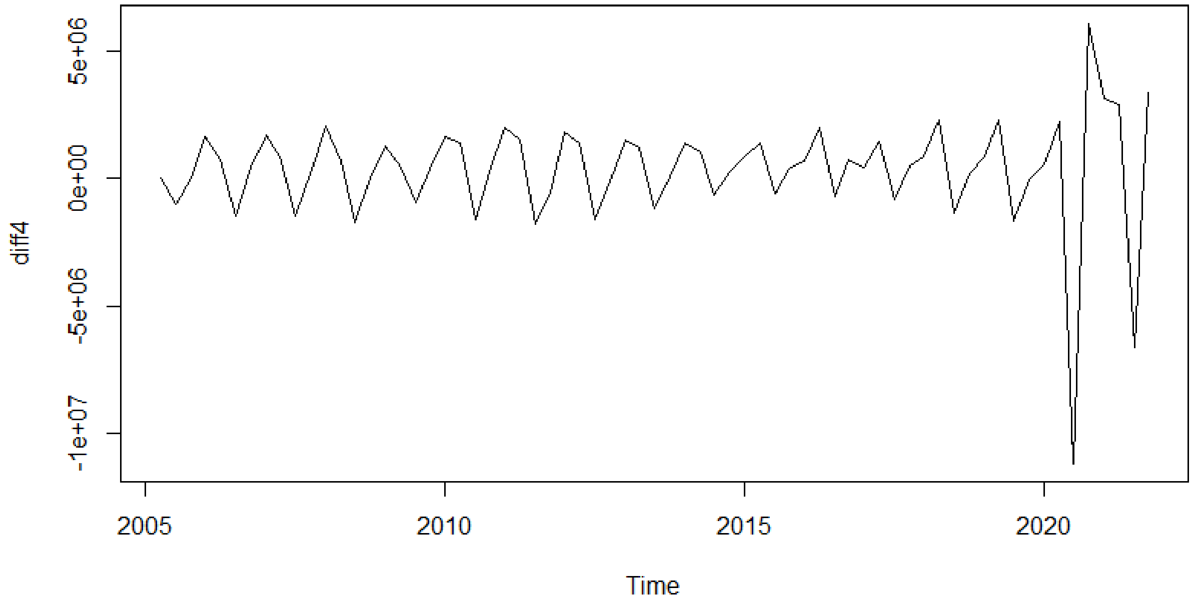
Dickey-Fuller = -2.5304, Lag order = 4, p-value = 0.3597

alternative hypothesis: stationary

#To Make Stationary

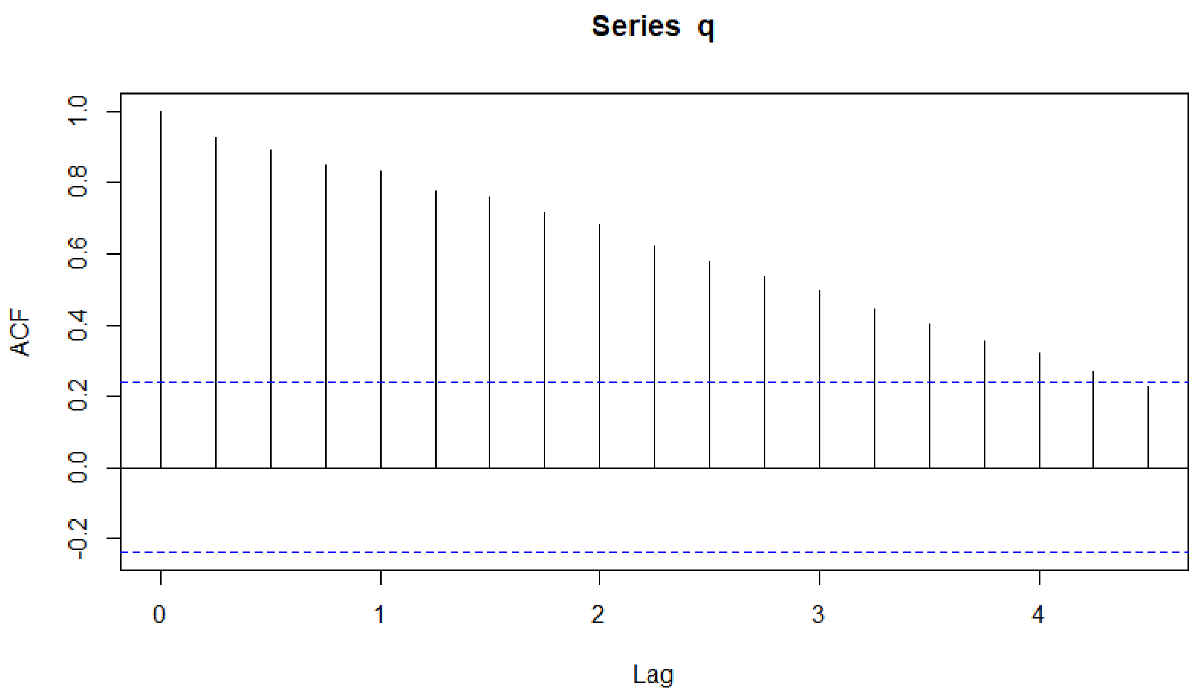
diff4<-diff(q)

plot(diff4)



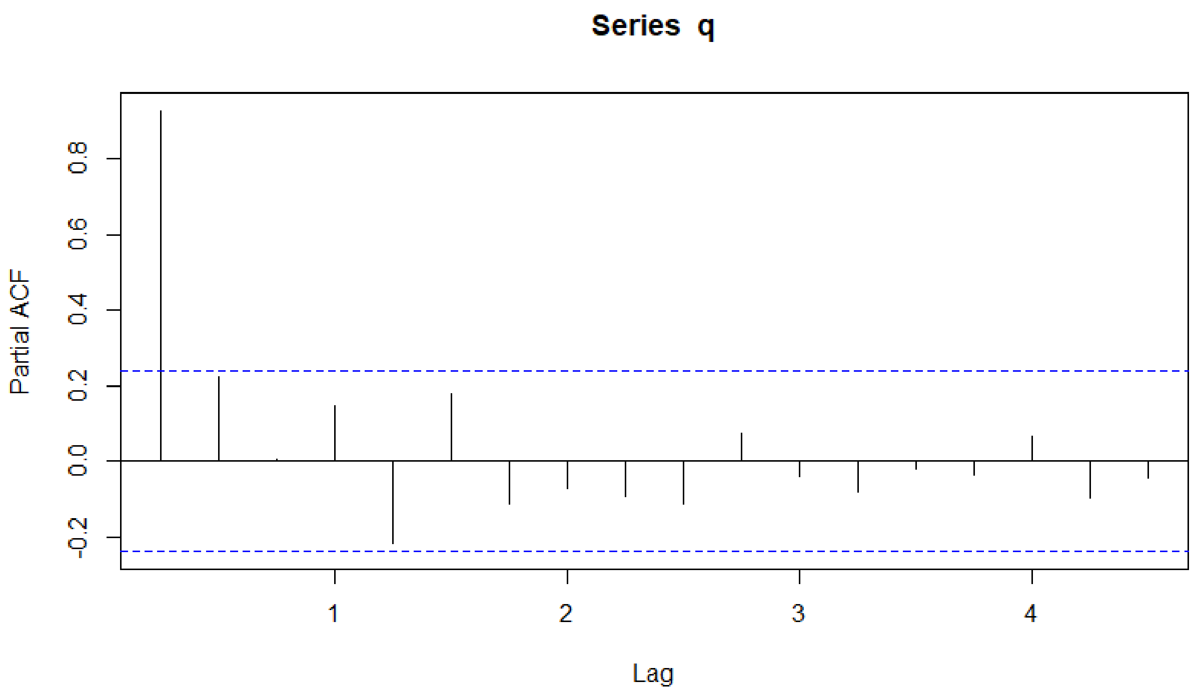
5e+06
0e+00
-5e+06
-1e+07
diff4
2005
2010
2015
2020
Time

acf(q)

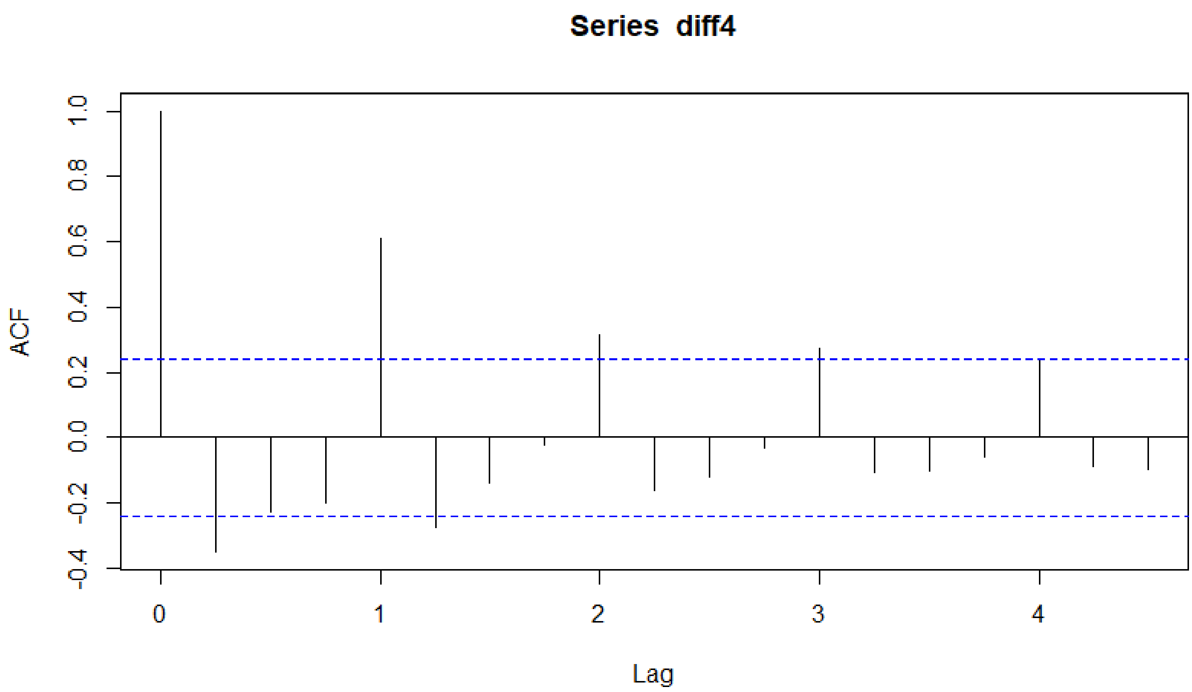


Series q
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
ACF
0
1
2
3
4
Lag

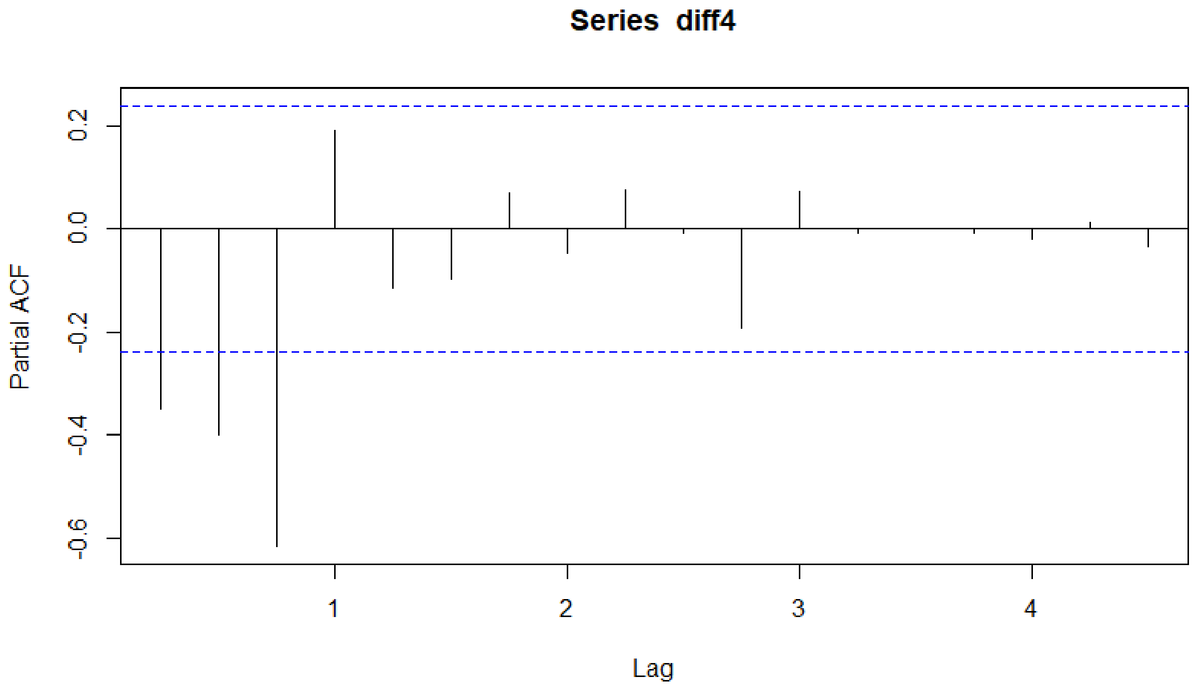
pacf(q)



acf(diff4)



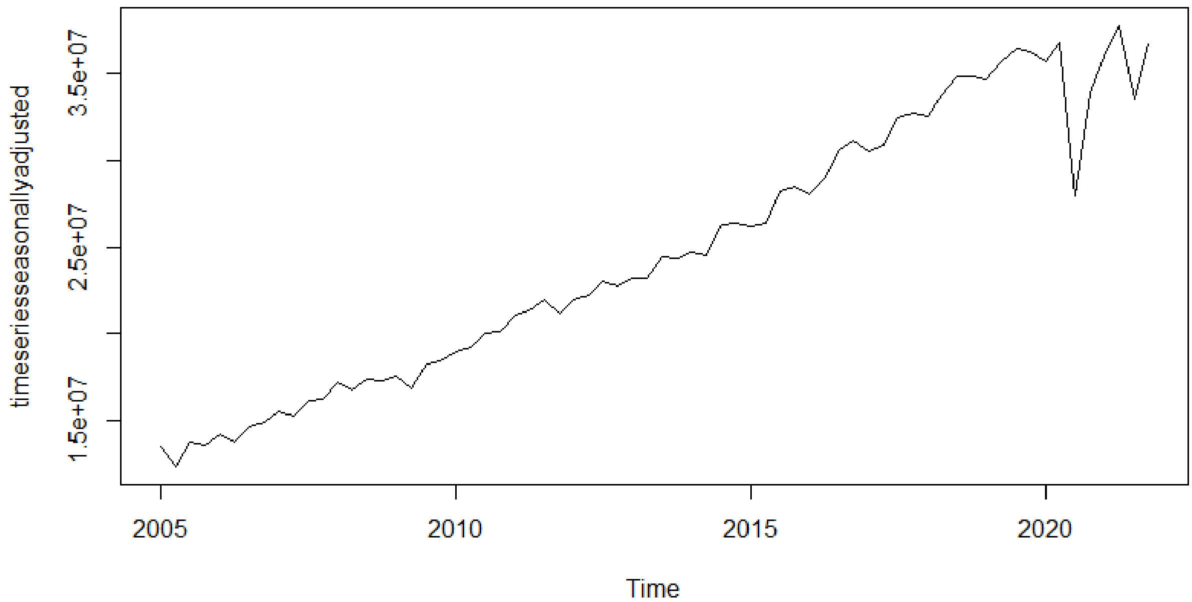
pacf(diff4)



#Remove Seasonality

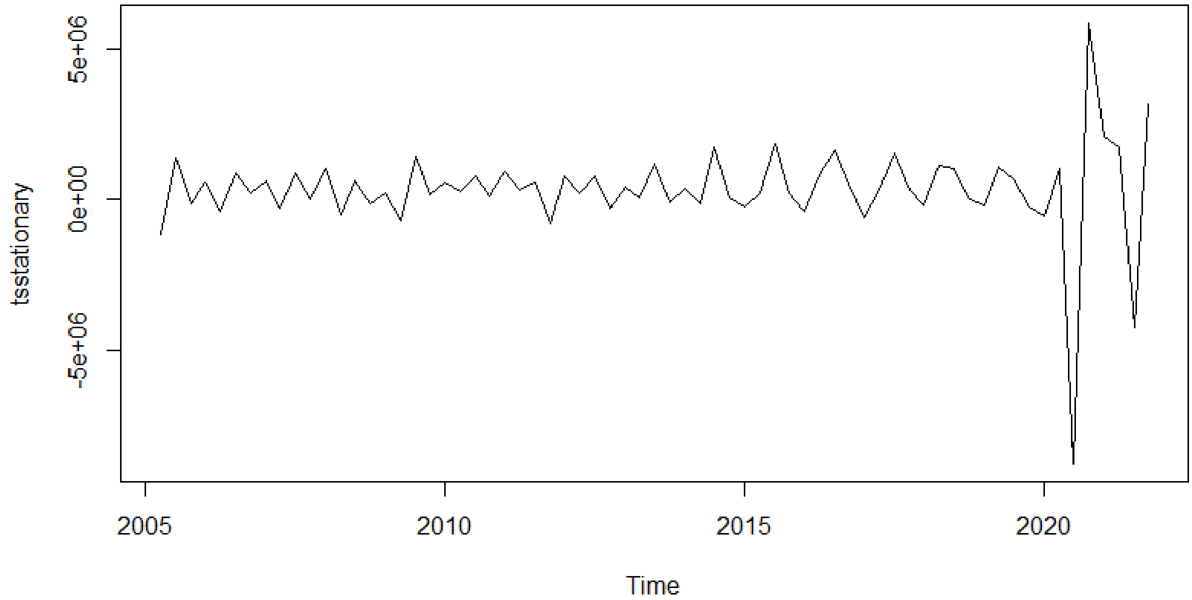
timeseriesseasonallyadjusted<-q-timeseriescomponent$seasonal

plot(timeseriesseasonallyadjusted)

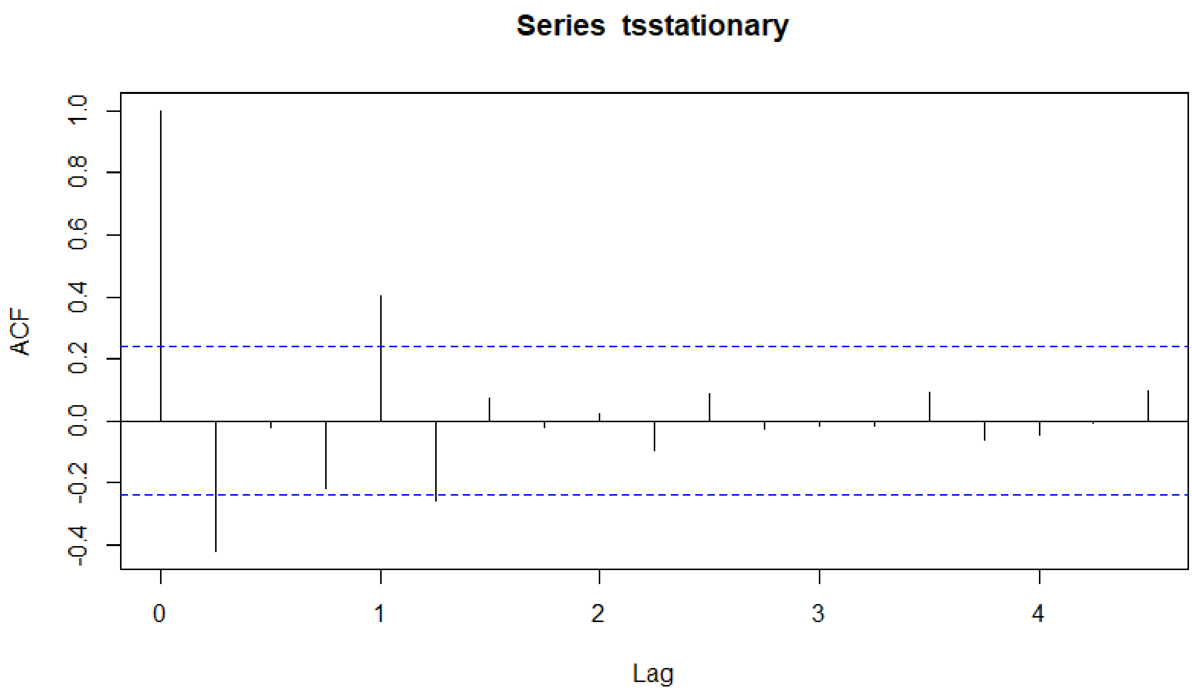


tsstationary<-diff(timeseriesseasonallyadjusted,differences = 1) plot(tsstationary)

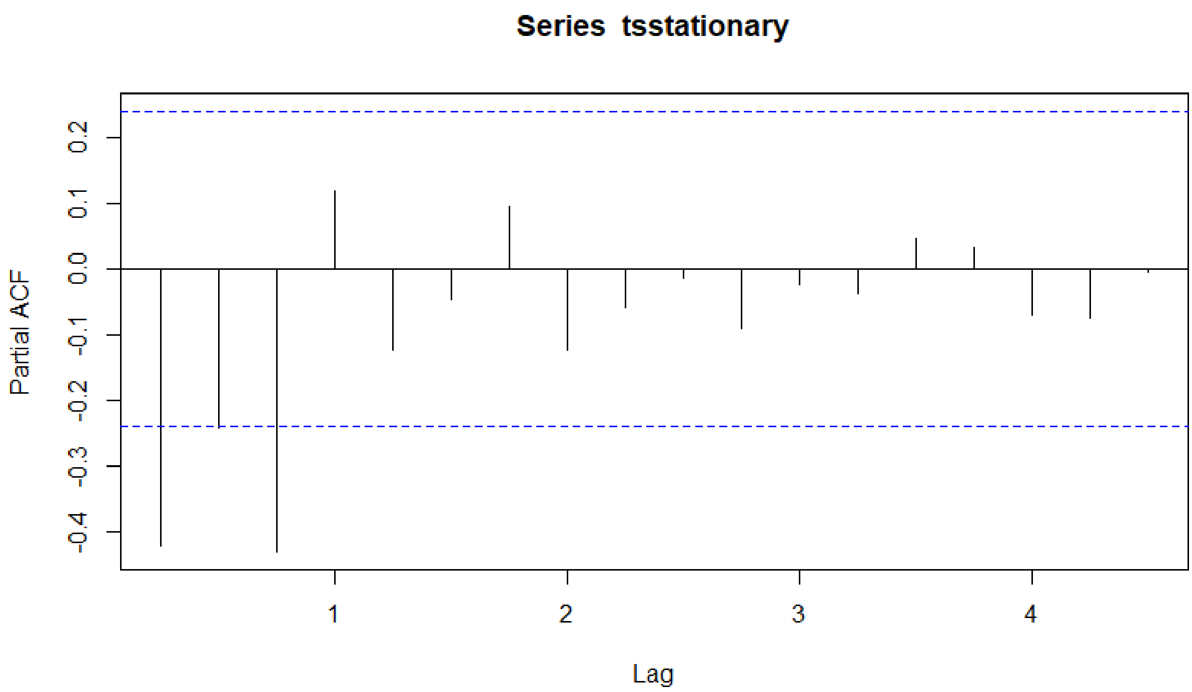
FIGURE - 5



acf(tsstationary)



pacf(tsstationary)



#Fit The Model

gdpqmodel=auto.arima(q)

gdpqmodel

Series: q

ARIMA(1,0,0)(0,1,1)[4] with drift

Coefficients:

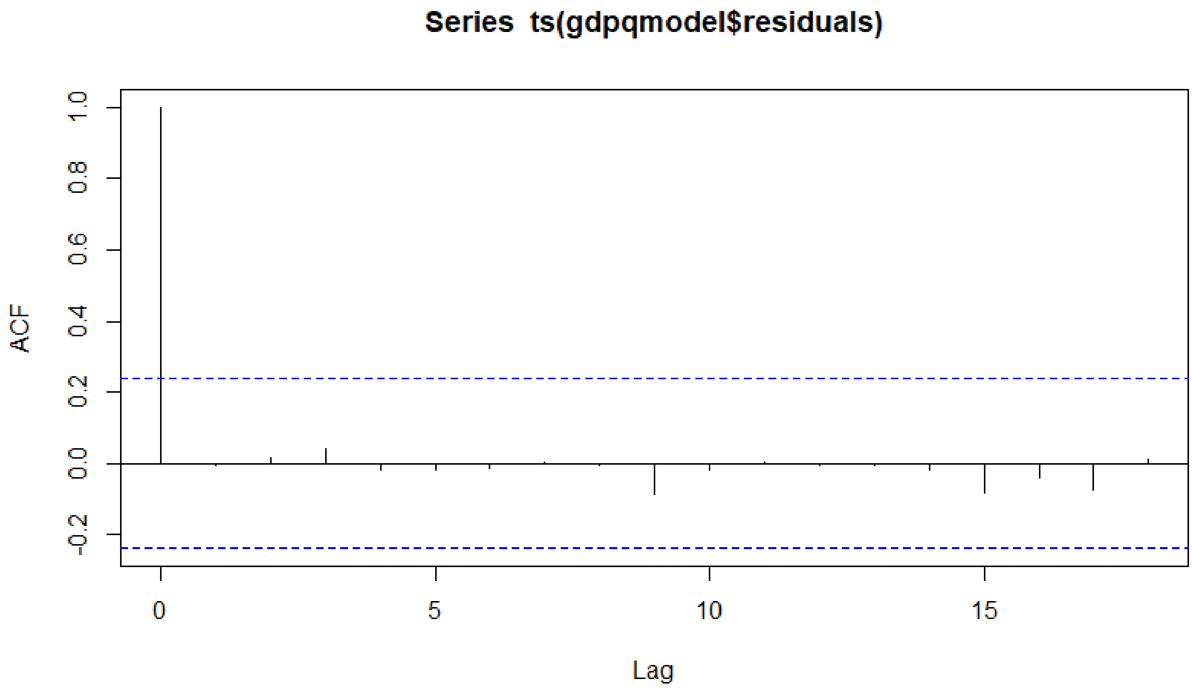
	ar1	sma1	drift
	0.4574	-0.3762	355982.34
s.e.	0.1106	0.1231	47787.38

sigma^2 = 1.758e+12: log likelihood = -991.92

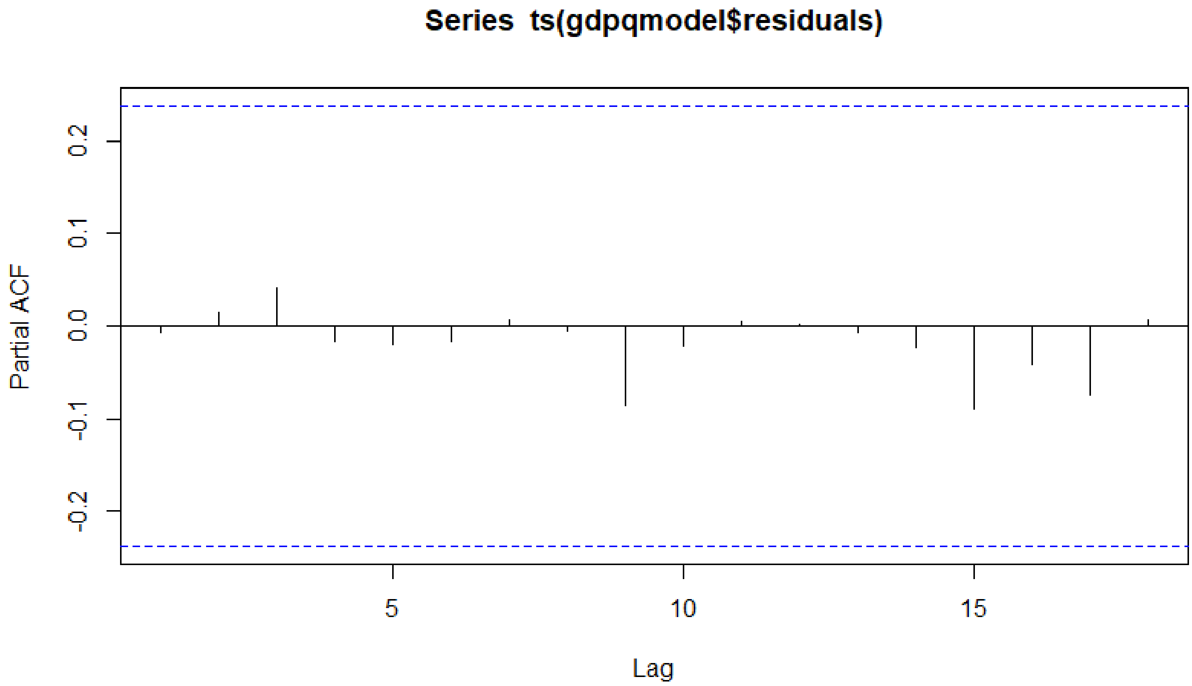
AIC=1991.85   AICc=1992.53   BIC=2000.48

#Residuals Diagnosis

acf(ts(gdpqmodel$residuals))



pacf(ts(gdpqmodel$residuals))



#Forecast Future Values

qforecast=forecast(gdpqmodel,level = c(95,80,90),h=5*4)

qforecast

	Point Forecast	Lo 80	Hi 80	Lo 90	Hi 90	Lo 95	Hi 95
2022 Q1	38118898	36419800	39817995	35938129	40299666	35520352	40717444
2022 Q2	40762825	38894396	42631253	38364723	43160927	37905310	43620340
2022 Q3	33784610	31882655	35686565	31343477	36225742	30875821	36693399
2022 Q4	37336242	35427346	39245137	34886201	39786283	34416838	40255646
2023 Q1	39540911	37320435	41761388	36690961	42390862	36144986	42936837
2023 Q2	42185878	39905583	44466173	39259151	45112604	38698468	45673288

2023 Q3 35208138 32915524 37500753 32265600 38150677 31701887 38714389

2023 Q4 38759988 36464804 41055171 35814151 41705824 35249807 42270168

2024 Q1 40964757 38415441 43514073 37692746 44236768 37065915 44863599

2024 Q2 43609769 41010418 46209120 40273538 46946000 39634404 47585134

2024 Q3 36632050 34022350 39241750 33282537 39981564 32640859 40623242

2024 Q4 40183909 37572049 42795769 36831623 43536195 36189413 44178404

2025 Q1 42388683 39551333 45226032 38746984 46030382 38049331 46728035

2025 Q2 45033697 42151394 47915999 41334301 48733092 40625595 49441798

2025 Q3 38055979 35164358 40947599 34344624 41767334 33633627 42478331

2025 Q4 41607838 38714271 44501405 37893985 45321690 37182510 46033166

2026 Q1 43812612 40714005 46911218 39835593 47789630 39073702 48551522

2026 Q2 46457626 43317808 49597443 42427713 50487538 41655689 51259563

2026 Q3 39479908 36331535 42628281 35439015 43520801 34664887 44294929

2026 Q4 43031767 39881607 46181927 38988581 47074954 38214013 47849521 summary(qforecast)

Forecast method: ARIMA(1,0,0)(0,1,1)[4] with drift

Model Information:

Series: q

ARIMA(1,0,0)(0,1,1)[4] with drift

Coefficients:

	ar1	sma1	drift
	0.4574	-0.3762	355982.34
s.e.	0.1106	0.1231	47787.38

sigma^2 = 1.758e+12: log likelihood = -991.92

AIC=1991.85 AICc=1992.53 BIC=2000.48

Error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-1101.739	1255720	490732.7	-0.2180837	1.946794	0.2799825	-0.007376611

Forecasts:

Point Forecast Lo 80 Hi 80 Lo 90 Hi 90 Lo 95 Hi 95

2022 Q1 38118898 36419800 39817995 35938129 40299666 35520352 40717444

2022 Q2 40762825 38894396 42631253 38364723 43160927 37905310 43620340

2022 Q3 33784610 31882655 35686565 31343477 36225742 30875821 36693399

2022 Q4 37336242 35427346 39245137 34886201 39786283 34416838 40255646

2023 Q1 39540911 37320435 41761388 36690961 42390862 36144986 42936837

2023 Q2 42185878 39905583 44466173 39259151 45112604 38698468 45673288

2023 Q3 35208138 32915524 37500753 32265600 38150677 31701887 38714389

2023 Q4 38759988 36464804 41055171 35814151 41705824 35249807 42270168

2024 Q1 40964757 38415441 43514073 37692746 44236768 37065915 44863599

2024 Q2 43609769 41010418 46209120 40273538 46946000 39634404 47585134

2024 Q3 36632050 34022350 39241750 33282537 39981564 32640859 40623242

2024 Q4 40183909 37572049 42795769 36831623 43536195 36189413 44178404

2025 Q1 42388683 39551333 45226032 38746984 46030382 38049331 46728035

2025 Q2 45033697 42151394 47915999 41334301 48733092 40625595 49441798

2025 Q3 38055979 35164358 40947599 34344624 41767334 33633627
42478331

2025 Q4 41607838 38714271 44501405 37893985 45321690 37182510
46033166

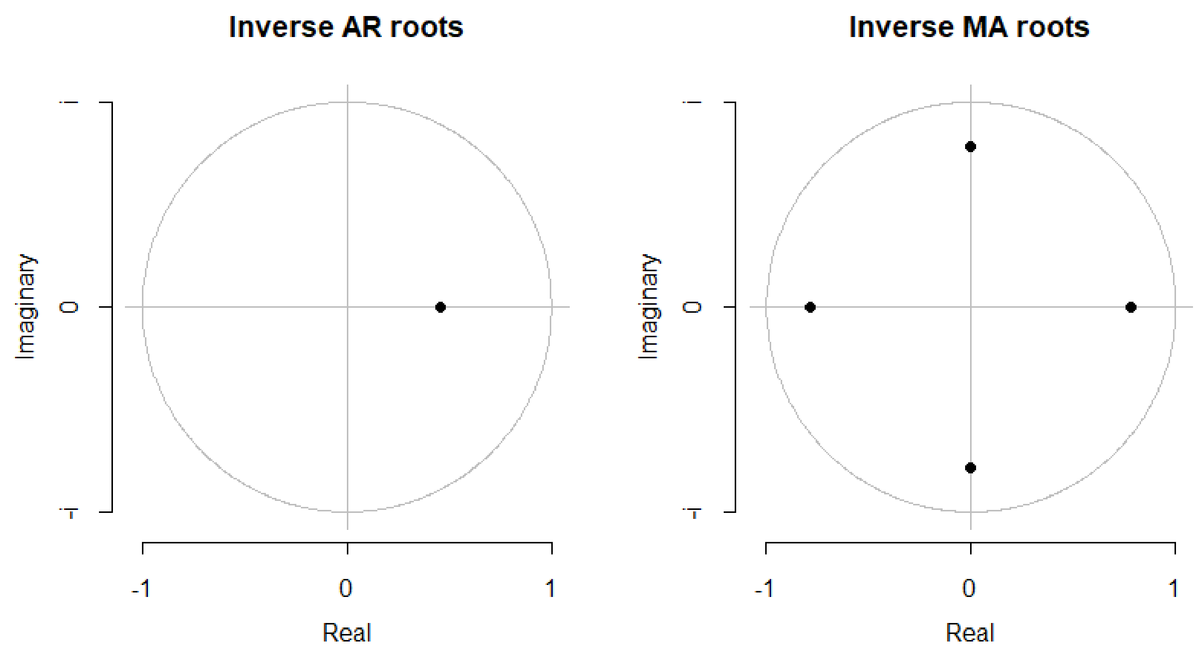
2026 Q1 43812612 40714005 46911218 39835593 47789630 39073702
48551522

2026 Q2 46457626 43317808 49597443 42427713 50487538 41655689
51259563

2026 Q3 39479908 36331535 42628281 35439015 43520801 34664887
44294929

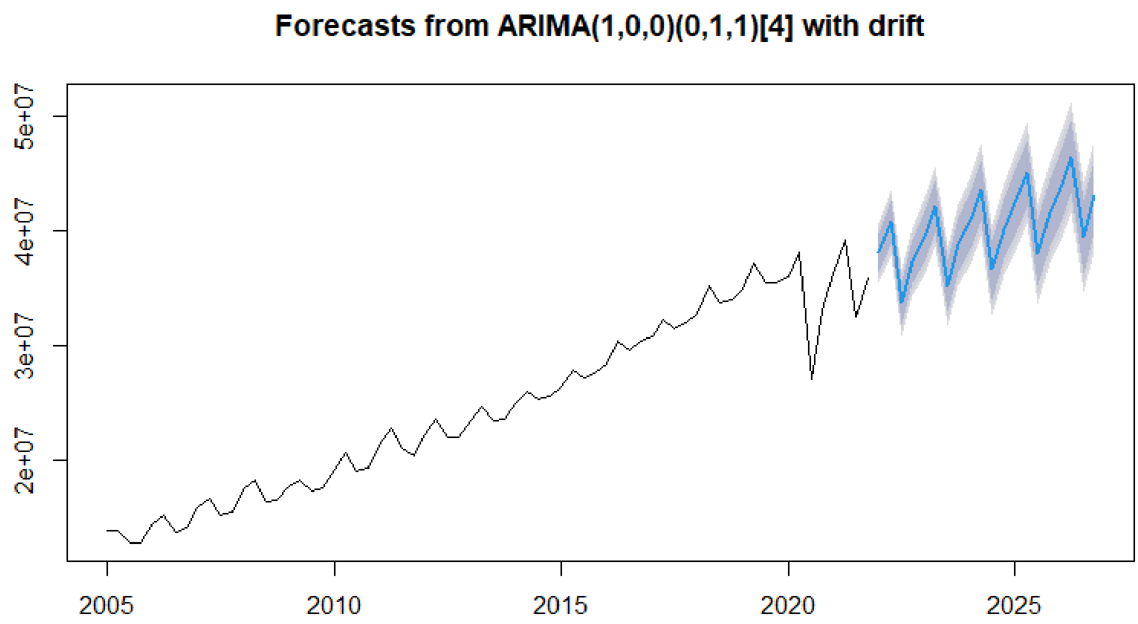
2026 Q4 43031767 39881607 46181927 38988581 47074954 38214013
47849521

plot(gdpqmodel)



plot(qforecast)

FIGURE -6



4.2) RESULTS

Figure 1 visualizes the GDP Growth in India from 1951 to 2020. The GDP growth has seen positive growth Annually.

The ACF and PACF plots were used to find the values of ARIMA parameters p,q,d . The ARIMA(1,2,1) were founded as the best model for forecasting.

ACF are the coefficients of correlation determine the relationship between a time series and its lags and 46 Using Time Series Forecasting for Analysis of GDP Growth in India .

PACF determines the partial correlation coefficients for ARIMA models.

ACF and PACF are used to indicate the relationship between the observations in time series and its lags for the time series and for determining the order parameters for ARIMA model.

Figure 3 displays the Forecast plot for GDP growth in India from 2021-2030 using ARIMA(1,2,1) model based on the Indian GDP values of last 70 years i.e. from 1951-2021

Figure 4 visualizes the GDP growth in India from 2005 to 2020(Quarterly data). The GDP Growth is unsteady and has seen Positive and Negative Growth.

Figure 6 plots the forecasted GDP Growth in India from 2021 to 2024.

5). CONCLUSION

In this study, we are using Time Series Forecasting Techniques ARIMA model to forecast GDP growth from 2021-2030.The Arima (1, 2, 1) fits the model and has indicated good predictive capability as per validation and performance parameters like RMSE, MAPE and MASE.The GDP forecast for the next decade is increasing over the period of time.Although the GDP of the country can be influenced by different conditions and thus predictions can vary in the future. This model has some limitations and future prospects can take into account other parameters affecting the economic conditions of the country as well. The monetary policy decisions by the policy makers make better use of forecasting macroeconomic variables such as GDP along with the inflation rate to assess the future state of economics. The proposed study gives the GDP forecasting and inflation rate can be future prospective of the given study.

6).References

1. https://tradingeconomics.com/india/gdp-growth-annual#:~:text=GDP%20Annual%20Growth%20Rate%20in,the%20second%20quarter%20of%202020
2. https://www.abs.gov.au/websitedbs/d3310114.nsf/home/time+series+analysis:+the+basics#:~:text=A%20time%20series%20is%20a,would%20comprise%20a%20time%20series
3. https://otexts.com/fpp2/arima.html
4. https://datascienceplus.com/time-series-analysis-using-arima-model-in-r/#:~:text=arima()%20function%20in%20R,chosen%20by%20minimizing%20the%20AICc
5. https://www.investopedia.com/terms/g/gdp.asp
6. https://fred.stlouisfed.org/
7. https://www.indiabudget.gov.in/economicsurvey/doc/Statistical-Appendix-in-English.pdf