Create Arima Model Gross domestic product (GDP) at Current Price Agriculture, Forestry & Fishing , Quarterly 2010 – 2022

By

Abuzar Yousef Ali Ahmed
Department of Basic Science, Faculty of preparatory Years
King Saud University

Doi: 10.21608/ajsr.2023.296407

Receiving the search 2 – 9 - 2022
Acceptance of publication 30 – 9 - 2022

Ahmed, Abuzar Yousef Ali (2023). Create Arima Model Gross domestic product (GDP) at Current Price Agriculture, Forestry & Fishing , Quarterly 2010 – 2022. ***The Arab Journal of Scientific Research***, AIESA, Egypt, 7(6), 41-70.

https://ajsr.journals.ekb.eg

Create Arima Model Gross domestic product (GDP) at Current Price Agriculture, Forestry & Fishing , Quarterly 2010 – 2022

Abstract

In this research, use the time series models to Create Arima Model Gross domestic. product (GDP) at current price Agriculture, Forestry & Fishing , Quarterly 2010 – 2022. The results showed that the model is the appropriate model for the series of Arima. Model Gross domestic product (GDP) at current price Agriculture, Forestry & Fishing, Quarterly is: ARIMA (3,1,1). According to the estimation results of this model, we observe the compatibility, between observed and estimated values as these values are consistent with those in the original time series, indicating the strength of the model and predictability.

Key words: Forecasting; MAPE; ARIMA models; stationary; identification ; estimation; MAE

Introduction

What can be forecast?

Forecasting is required in many situations: deciding whether to build another power generation plant in the next five years requires forecasts of future demand; scheduling staff in a call centre next week requires forecasts of call volumes; stocking an inventory requires forecasts of stock requirements. Forecasts can be required several years in advance (for the case of capital investments), or only a few minutes beforehand (for telecommunication routing). Whatever the circumstances or time horizons involved, forecasting is an important aid to effective and efficient planning.

Some things are easier to forecast than others. The time of the sunrise tomorrow morning can be forecast precisely. On the other hand, tomorrow's lotto numbers cannot be forecast with any accuracy. The predictability of an event or a quantity depends on several factors including:

1. how well we understand the factors that contribute to it؛
2. how much data is available؛
3. whether the forecasts can affect the thing we are trying to forecast.

For example, forecasts of electricity demand can be highly accurate because all three conditions are usually satisfied. We have a good idea of the contributing factors:

- electricity demand is driven largely by temperatures, with smaller effects for calendar
- variation such as holidays, and economic conditions. Provided there is a sufficient
- history of data on electricity demand and weather conditions, and we have the skills to
- develop a good model linking electricity demand and the key driver variables, the
- forecasts can be remarkably accurate.

A lot of resources for time series analysis are available to the R community including:

 several useful individual functions (such as plotting the sample autocorrelation and sample partial autocorrelation functions, fitting an ARIMA Model etc. for regularly spaced time series) included with the base R infrastructure.

 additional packages for more extensive time series analysis, and for state-space models and spectral analysis.

 time series datasets available directly in base R and in other time series packages

 books, on-line tutorials, and other on-line resources

Forecasting, planning and goals

Forecasting is a common statistical task in business, where it helps to inform decisions about the scheduling of production, transportation and personnel, and provides a guide to

ISSN: 2537-0367 eISSN : 2537-0375

long-term strategic planning. However, business forecasting is often done poorly, and is frequently confused with planning and goals. They are three different things.

Forecasting is about predicting the future as accurately as possible, given all of the information available, including historical data and knowledge of any future events that might impact the forecasts.

Goals are what you would like to have happen. Goals should be linked to forecasts and plans, but this does not always occur. Too often, goals are set without any plan for how to achieve them, and no forecasts for whether they are realistic.

Planning is a response to forecasts and goals. Planning involves determining the appropriate actions that are required to make your forecasts match your goals.

Forecasting should be an integral part of the decision-making activities of management, as it can play an important role in many areas of a company. Modern organizations require short-term, medium-term and long-term forecasts, depending on the specific application.

Short-term forecasts are needed for the scheduling of personnel, production and transportation. As part of the scheduling process, forecasts of demand are often also required.

Medium-term forecasts are needed to determine future resource requirements, in order to purchase raw materials, hire personnel, or buy machinery and equipment.

Long-term forecasts are used in strategic planning. Such decisions must take account of market opportunities, environmental factors and internal resources.

An organization needs to develop a forecasting system that involves several approaches to predicting uncertain events. Such forecasting systems require the development of expertise in identifying forecasting problems, applying a range of forecasting methods, selecting appropriate methods for each problem, and

ISSN: 2537-0367 eISSN : 2537-0375

evaluating and refining forecasting methods over time. It is also important to have strong organizational support for the use of formal forecasting methods if they are to be used successfully.

Determining what to forecast In the early stages of a forecasting project, decisions need to be made about what should be forecast. For example, if forecasts are required for items in a manufacturing environment, it is necessary to ask whether forecasts are needed for:

1. every product line, or for groups of products?
2. every sales outlet, or for outlets grouped by region, or only for total sales؟
3. weekly data, monthly data or annual data?
4. It is also necessary to consider the forecasting horizon. Will forecasts be required for one month in advance, for 6 months, or for ten years? Different types of models will be necessary, depending on what forecast horizon is most important.

How frequently are forecasts required? Forecasts that need to be produced frequently are better done using an automated system than with methods that require careful manual work.

It is worth spending time talking to the people who will use the forecasts to ensure that you understand their needs, and how the forecasts are to be used, before embarking on extensive work in producing the forecasts.

Once it has been determined what forecasts are required, it is then necessary to find or collect the data on which the forecasts will be based. The data required for forecasting may already exist. These days, a lot of data are recorded, and the forecaster's task is often to identify where and how the required data are stored. The data may include sales records of a company, the historical demand for a product, or the unemployment rate for a geographic region. A large part of a

ISSN: 2537-0367 **eISSN : 2537-0375**

forecaster's time can be spent in locating and collating the available data prior to developing suitable forecasting methods.

Forecasting data and methods The appropriate forecasting methods depend largely on what data are available. If there are no data available, or if the data available are not relevant to the forecasts. then qualitative forecasting methods must be used. These methods are not purely guesswork—there are well-developed structured approaches to obtaining good forecasts without using historical data. These methods are discussed in Chapter 4.

Quantitative forecasting can be applied when two conditions are satisfied:

1. numerical information about the past is available؛
2. it is reasonable to assume that some aspects of the past patterns will continue into the future.

There is a wide range of quantitative forecasting methods, often developed within specific disciplines for specific purposes. Each method has its own properties. accuracies, and costs that must be considered when choosing a specific method.

Most quantitative prediction problems use either time series data (collected at regular intervals over time) or cross-sectional data (collected at a single point in time). In this book we are concerned with forecasting future data, and we concentrate on the time series domain.

Time series forecasting

Time series analysis is a statistical method for analyzing and modeling the trends and patterns in a series of data points collected over time. It is commonly used in fields such as economics, finance, and environmental science to study patterns in data and make predictions about future events.

There are several techniques used in time series analysis, including:

ISSN: 2537-0367 eISSN : 2537-0375

- Decomposition: This involves breaking down the time series into its component parts. such as trend, seasonality, and residuals.
- Smoothing: This involves using mathematical techniques to smooth out the fluctuations in the time series data to better visualize trends and patterns.
- ARIMA modeling: This involves fitting an autoregressive integrated moving average (ARIMA) model to the time series data to make predictions about future events.
- Exponential smoothing: This is a popular method for forecasting future values in a time series by using a weighted average of past observations.
- Fourier Analysis: This involves transforming a time series into the frequency domain to identify the frequencies of different components in the data.
- Time series analysis is a powerful tool for understanding patterns in data and making predictions, but it can also be complex and challenging.

It's important to carefully select the appropriate method and interpret the results in light of the underlying assumptions and limitations of the chosen approach.

es, that's correct! Time series analysis is a powerful tool, but it is also important to understand the assumptions and limitations of each technique in order to interpret the results correctly.

Additionally, it's also important to make sure that the data being analyzed is stationary, meaning that its mean and variance are constant over time. If the data is not stationary, it can be made stationary through techniques such as differences or transformations, such as the log transformation.

Another important aspect of time series analysis is to check for any outliers or anomalies in the data, as these can have

ISSN: 2537-0367 **eISSN : 2537-0375**

a significant impact on the results. Outliers can be detected using statistical techniques such as the Z-score or the median absolute deviation.

In conclusion, time series analysis is a valuable tool for understanding patterns in data over time, but it is also important to be mindful of its limitations and assumptions, as well as to take steps to pre-process the data before conducting the analysis.

Examples of time series data include:

- Daily IBM stock prices
- Monthly rainfall
- Quarterly sales results for Amazon
- Annual Google profits

Anything that is observed sequentially over time is a time series. we will only consider time series that are observed at regular intervals of time (e.g., hourly, daily, weekly,monthly, quarterly, annually). Irregularly spaced time series can also occur.

When forecasting time series data, the aim is to estimate how the sequence of observations will continue into the future. Figure 1.1 shows the quarterly Australian beer production from 1992 to the second quarter of 2010.

Figure 1.1: Australian quarterly beer production: 1992Q1–2010Q2, with two years of forecasts.

The blue lines show forecasts for the next two years. Notice how the forecasts have captured the seasonal pattern seen in the historical data and replicated it for the next two years. The dark shaded region shows 80% prediction intervals. That is, each future value is expected to lie in the dark shaded region with a probability of 80%. The light shaded region shows 95% prediction intervals. These prediction intervals are a useful way of displaying the uncertainty in forecasts. In this case the forecasts are expected to be accurate, and hence the prediction intervals are quite narrow.

ISSN: 2537-0367 eISSN : 2537-0375

The simplest time series forecasting methods use only information on the variable to be forecast, and make no attempt to discover the factors that affect its behavior.

Therefore they will extrapolate trend and seasonal patterns, but they ignore all other information such as marketing initiatives, competitor activity, changes in economic conditions, and so on.

Time series models used for forecasting include decomposition models, exponential smoothing models and ARIMA models.

Predictor variables and time series forecasting

Predictor variables are often useful in time series forecasting. For example, suppose we wish to forecast the hourly electricity demand (ED) of a hot region during the summer period. A model with predictor variables might be of the form ED=f(current) temperature, strength of economy, population, time of day, day of week, (error).ED=f(current temperature, strength of economy, population, time of day, day of week, error).The relationship is not exact — there will always be changes in electricity demand that cannot be accounted for by the predictor variables. The “error” term on the right allows for random variation and the effects of relevant variables that are not included in the model. We call this an explanatory model because it helps explain what causes the variation in electricity demand.

Because the electricity demand data form a time series, we could also use a time series model for forecasting. In this case, a suitable time series forecasting equation is of the formEDt+1=f(EDt,EDt−1,EDt−2,EDt−3,…,error),EDt+1=f(EDt, EDt−1,EDt−2,EDt−3,…,e
rror),where tt is the present hour, t+1t+1 is the next hour, t−1t−1 is the previous

ISSN: 2537-0367 eISSN : 2537-0375

hour, t−2t−2 is two hours ago, and so on. Here, prediction of the future is based on past values of a variable, but not on external variables which may affect the system.

Again, the “error” term on the right allows for random variation and the effects of relevant variables that are not included in the model.

There is also a third type of model which combines the features of the above two models. For example, it might be given byEDt+1=f(EDt, current temperature, time of day, day of week, error).EDt+1=f(EDt,current temperature, time of day, day of week, (error).These types of mixed models have been given various names in different disciplines. They are known as dynamic regression models, panel data models, longitudinal models, transfer function models, and linear system models (assuming that ff is linear).

An explanatory model is useful because it incorporates information about other variables, rather than only historical values of the variable to be forecast. However, there are several reasons a forecaster might select a time series model rather than an explanatory or mixed model. First, the system may not be understood, and even if it was understood it may be extremely difficult to measure the relationships that are assumed to govern its behaviour. Second, it is necessary to know or forecast the future values of the various predictors in order to be able to forecast the variable of interest, and this may be too difficult. Third, the main concern may be only to predict what will happen, not to know why it happens. Finally, the time series model may give more accurate forecasts than an explanatory or mixed model.

The model to be used in forecasting depends on the resources and data available, the accuracy of the competing models, and the way in which the forecasting model is to be used.

ISSN: 2537-0367 eISSN : 2537-0375

The basic steps in a forecasting task

A forecasting task usually involves five basic steps.

Step 1: Problem definition.

Often this is the most difficult part of forecasting. Defining the problem carefully requires an understanding of the way the forecasts will be used, who requires the forecasts, and how the forecasting function fits within the organization requiring the forecasts. A forecaster needs to spend time talking to everyone who will be involved in collecting data, maintaining databases, and using the forecasts for future planning.

Step 2: Gathering information.

There are always at least two kinds of information required: (a) statistical data, and (b) the accumulated expertise of the people who collect the data and use the forecasts. Often, it will be difficult to obtain enough historical data to be able to fit a good statistical model. In that case, old data will be less useful due to structural changes in the system being forecast; then we may choose to use only the most recent data. However, remember that good statistical models will handle evolutionary changes in the system; don’t throw away good data unnecessarily.

Step 3: Preliminary (exploratory) analysis.

Always start by graphing the data. Are there consistent patterns? Is there a significant trend? Is seasonality important? Is there evidence of the presence of business cycles? Are there any outliers in the data that need to be explained by those with expert knowledge? How strong are the relationships among the variables available for analysis? Various tools have been developed to help with this analysis.

Step 4: Choosing and fitting models.

The best model to use depends on the availability of historical data, the strength of relationships between the forecast variable and any explanatory variables, and the way in which the

ISSN: 2537-0367 eISSN : 2537-0375

forecasts are to be used. It is common to compare two or three potential models. Each model is itself an artificial construct that is based on a set of assumptions (explicit and implicit) and usually involves one or more parameters which must be estimated using the known historical data

Step 5: Using and evaluating a forecasting model.

Once a model has been selected and its parameters estimated, the model is used to make forecasts. The performance of the model can only be properly evaluated after the data for the forecast period have become available. A number of methods have been developed to help in assessing the accuracy of forecasts. There are also organizational issues in using and acting on the forecasts. When using a forecasting model in practice, numerous practical issues arise such as how to handle missing values and outliers, or how to deal with short time series.

The statistical forecasting perspective

The thing we are trying to forecast is unknown (or we would not be forecasting it), and so we can think of it as a random variable. For example, the total sales for next month could take a range of possible values, and until we add up the actual sales at the end of the month, we don't know what the value will be. So until we know the sales for next month, it is a random quantity.

Because next month is relatively close, we usually have a good idea what the likely sales values could be. On the other hand, if we are forecasting the sales for the same month next year, the possible values it could take are much more variable. In most forecasting situations, the variation associated with the thing we are forecasting will shrink as the event approaches. In other words, the further ahead we forecast, the more uncertain we are.

We can imagine many possible futures, each yielding a different value for the thing we wish to forecast. Plotted in black

ISSN: 2537-0367 eISSN : 2537-0375

in Figure 1.2 are the total international visitors to Australia from 1980 to 2015. Also shown are ten possible futures from 2016–2025.

Figure 1.2: Total international visitors to Australia (1980-2015) along with ten possible futures.

When we obtain a forecast, we are estimating the middle of the range of possible values the random variable could take. Often, a forecast is accompanied by a prediction interval giving a range of values the random variable could take with relatively high probability. For example, a 95% prediction interval contains a range of values which should include the actual future value with probability 95%. Rather than plotting individual possible futures as shown in Figure 1.2, we usually show these prediction intervals instead. The plot below shows 80% and 95% intervals for the future Australian international visitors. The blue line is the average of the possible future values, which we call the point forecasts.

Figure 1.3: Total international visitors to Australia (1980–2015) along with 10-year
forecasts and 80% and 95% prediction intervals.

We will use the subscript tt for time. For example, ytyt will denote the observation at time tt. Suppose we denote all the information we have observed as II and we want to forecast ytyt. We then write yt|Iyt|I meaning "the random variable ytyt given what we know in II". The set of values that this random variable could take, along with their relative probabilities, is known as the "probability distribution" of yt|Iyt|I. In forecasting, we call this the forecast distribution.

When we talk about the "forecast", we usually mean the average value of the forecast distribution, and we put a "hat" over yy to show this. Thus, we write the forecast of ytyt as ^yty^t, meaning the average of the possible values that ytyt could take given

ISSN: 2537-0367 eISSN : 2537-0375

everything we know. Occasionally, we will use ^yty^t to refer to the median (or middle value) of the forecast distribution instead.
It is often useful to specify exactly what information we have used in calculating the forecast. Then we will write, for example, ^yt|t−1y^t|t−1 to mean the forecast of ytyt taking account of all previous observations (y1,…,yt−1)(y1,…,yt−1.(Similarly, ^yT+h|Ty^T+h|T means the forecast of yT+hyT+h taking account of y1,…,yTy1,…,yT (i.e., an hh-step forecast taking account of all observations up to time TT.

Basics of Time-Series Forecasting

Timeseries forecasting in simple words means to forecast or to predict the future value(eg-stock price) over a period of time. There are different approaches to predict the value, consider an example there is a company XYZ records the website traffic in each hour and now wants to forecast the total traffic of the coming hour. If I ask you what will your approach to forecasting the upcoming hour traffic? A different person can have a different perspective like one can say find the mean of all observations, one can have like take mean of recent two observations, one can say like give more weightage to current observation and less to past, or one can say use interpolation. There are different methods to forecast the values.

while Forecasting time series values, 3 important terms need to be taken care of and the main task of time series forecasting is to forecast these three terms.

Seasonality

Seasonality is a simple term that means while predicting a time series data there are some months in a particular domain where the output value is at a peak as compared to other months. for example if you observe the data of tours and travels companies of past 3 years then you can see that in November and December the distribution will be very high due to holiday

ISSN: 2537-0367 eISSN : 2537-0375

season and festival season. So while forecasting time series data we need to capture this seasonality.

Trend

The trend is also one of the important factors which describe that there is certainly increasing or decreasing trend time series, which actually means the value of organization or sales over a period of time and seasonality is increasing or decreasing.

Unexpected Events

Unexpected events mean some dynamic changes occur in an organization, or in the market which cannot be captured. for example a current pandemic we are suffering from, and if you observe the Sensex or nifty chart there is a huge decrease in stock price which is an unexpected event that occurs in the surrounding.

Methods and algorithms are using which we can capture seasonality and trend But the unexpected event occurs dynamically so capturing this becomes very difficult.

Rolling Statistics and Stationarity in Time-series

A stationary time series is a data that has a constant mean and constant variance. If I
take a mean of T1 and T2 and compare it with the mean of T4 and T5 then is it the same, and if different, how much difference is there? So, constant mean means this difference should be less, and the same with variance.

If the time series is not stationary, we have to make it stationary and then proceed with modelling. Rolling statistics is help us in making time series stationary. so basically rolling statistics calculates moving average. To calculate the moving average we need to define the window size which is basically how much past values to be considered.

ISSN: 2537-0367 eISSN : 2537-0375

For example, if we take the window as 2 then to calculate a moving average in the above example then, at point T1 it will be blank, at point T2 it will be the mean of T1 and T2, at point T3 mean of T3 and T2, and so on. And after calculating all moving averages if you plot the line above actual values and calculated moving averages then you can see that the plot will be smooth.

This is one method of making time series stationary, there are other methods also which we are going to study as Exponential smoothing.

Additive and Multiplicative Time series

In the real world, we meet with different kinds of time series data. For this, we must know the concepts of Exponential smoothing and for this first, we need to study types of time series data as additive and multiplicative. As we studied there are 3 components we need to capture as Trend(T), seasonality(S), and Irregularity(I).

Additive time series is a combination(addition) of trend, seasonality, and Irregularity while multiplicative time series is the multiplication of these three terms.

Model Type. The following options are available:

- All models. The Expert Modeler considers both ARIMA and exponential smoothing models.
- Exponential smoothing models only. The Expert Modeler only considers exponential smoothing models.
- ARIMA models only. The Expert Modeler only considers ARIMA models.

Expert Modeler considers seasonal models. This option is only enabled if a periodicity has been defined for the active dataset. When this option is selected, the Expert Modeler considers both seasonal and nonseasonal models. If this option is not selected, the Expert Modeler only considers nonseasonal models.

ISSN: 2537-0367 eISSN : 2537-0375

Events and Interventions. Enables you to designate certain input fields as event or intervention fields. Doing so identifies a field as containing time series data affected by events (predictable recurring situations, for example, sales promotions) or interventions (one-time incidents, for example, power outage or employee strike). The Expert Modeler does not consider arbitrary transfer functions for inputs identified as event or intervention fields.

Input fields must have a measurement level of Flag, Nominal, or Ordinal and must be numeric (for example, 1/0, not True/False, for a flag field), before they will be included in this list.

Outliers

Detect outliers automatically. By default, automatic detection of outliers is not performed. Select this option to perform automatic detection of outliers, then select the desired outlier types. See the topic Handling Outliers for more information.

Related information:

Streaming TS Model Options Handling Outliers

Univariate Series (TSMODEL algorithms)

Users can let the Expert Modeler select a model for them from:

- All models (default).
- Exponential smoothing models only.
- ARIMA models only.

ARIMA Expert Model (TSMODEL algorithms)

Transformation (none, log or sqrt?)

Series

Seasonal length

Impute missing

Difference order

Pattern detection (ACF, PACF, EACF) for initial model

Fit the model by CLS

ISSN: 2537-0367 eISSN : 2537-0375

Fit the model by ML
Diagnostic checking
Ljung-Box, ACF/PACF
ARIMA EM
Delete insignificant
parameters in 3 stages:
|t|<0.5, 2. |t|<1, 3. |t|<2

1. Delete insignificant parametersModify model (only once)
2. Gross domestic product (GDP(

Gross domestic product (GDP) is the standard measure of the value added created through the production of goods and services in a country during a certain period. As such, it also measures the income earned from that production, or the total amount spent on final goods and services (less imports). While GDP is the single most important indicator to capture economic activity, it falls short of providing a suitable measure of people's material well-being for which alternative indicators may be more appropriate. This indicator is based on nominal GDP (also called GDP at current prices or GDP in value) and is available in different.

measures: US dollars and US dollars per capita (current PPPs). All OECD countries compile their data according to the 2008 System of National Accounts (SNA). This indicator is less suited for comparisons over time, as developments are not only caused by real growth, but also by changes in prices and PPPs.

Quarterly GDP

Gross domestic product (GDP) is the standard measure of the value added created through the production of goods and services in a country during a certain period.

As such, it also measures the income earned from that production, or the total amount spent on final goods and services (less imports). While GDP is the single most important indicator to capture economic activity, it falls short of providing a suitable

ISSN: 2537-0367 eISSN : 2537-0375

measure of people's material well-being for which alternative indicators may be more appropriate. This indicator is based on real GDP (also called GDP at constant prices or GDP in volume), i.e. the developments over time are adjusted for price changes. The numbers are also adjusted for seasonal influences. The indicator is available in different measures: percentage change from the previous quarter, percentage change from the same quarter of the previous year and volume index (2015=100). All OECD countries compile their data according to the 2008 System of National Accounts (SNA).

Real GDP forecast

Real gross domestic product (GDP) is GDP given in constant prices and refers to the volume level of GDP. Constant price estimates of GDP are obtained by expressing values of all goods and services produced in a given year, expressed in terms of a base period. Forecast is based on an assessment of the economic climate in individual countries and the world economy, using a combination of model-based analyses and expert judgement. This indicator is measured in growth rates compared to previous year.

Real GDP long-term forecast

Trend gross domestic product (GDP), including long-term baseline projections (upto 2060), in real terms. Forecast is based on an assessment of the economic climate in individual countries and the world economy, using a combination of model-based analyses and expert judgement. This indicator is measured in USD at constant prices and Purchasing Power Parities (PPPs)

Data Analysis

Time Series Modeler

Model Description

Model Type

The model description table contains an entry for each estimated model and includes both a model identifier and the

ISSN: 2537-0367 eISSN : 2537-0375

model type. The model identifier consists of the name)or label) of the associated dependent variable and a system-assigned name. In the current example, the dependent variable is Sales of Men's Clothing and the system-assigned name is Model_1.

The Time Series Modeler supports both exponential smoothing and ARIMA models.

Exponential smoothing model types are listed by their commonly used names such as Holt and Winters' Additive. ARIMA model types are listed using the standard notation of ARIMA(p,d,q) (P,D,Q), where p is the order of autoregression, d is the order of differencing (or integration), and q is the order of moving-average, and (P,D,Q) are their seasonal counterparts.

The Expert Modeler has determined that sales of men's clothing is best described by a seasonal ARIMA model with one order of differencing. The seasonal nature of the model accounts for the seasonal peaks that we saw in the series plot, and the single order of differencing reflects the upward trend that was evident in the data.

Model Summary

Model Fit

Fit

Statistic

Mean SE Minimum Maximum Percentile

Stationary

R– squared

887.887.887.887.887.887.887.887.887.887.

R– squared

958.985.985.985.985.985.985.985.985.

RMSE 741.081 . 741.081 741.081 741.081 741.081 741.081 741.081 741.081 741.081 741.081

MAPE 2.916 . 2.916 2.916 2.916 2.916 2.916 2.916 2.916 2.916 2.916

ISSN: 2537-0367 eISSN : 2537-0375

MaxAPE 22.158 . 22.158 22.158 22.158 22.158 22.158 22.158 22.158 22.158 22.158
MAE 491.095 . 491.095 491.095 491.095 491.095 491.095 491.095 491.095 491.095 491.095
MaxAE 2525.869 . 2525.869 2525.869 2525.869 2525.869 2525.869 2525.869 2525.869 2525.869 2525.869
Normalized
BIC
13.627 13.627 13.627 13.627 13.627 13.627 13.627
 R-squared
 RMSE (Root Mean Square Error)
 MAPE (Mean Absolute Percentage Error)
 MAE (Mean Absolute Error)
 MaxAPE (Maximum Absolute Percentage Error)
 MaxAE (Maximum Absolute Error)
 Norm. BIC (Normalized Bayesian Information Criterion)
Generate. Enables you to generate a Time Series modeling node back to the stream
or a model nugget to the palette.
 Generate Modeling Node. Places a Time Series modeling node into a stream
with the settings used to create this set of models. Doing so would be useful, for
example, if you have a stream in which you want to use these model settings
but you no longer have the modeling node used to generate them.
 Model to Palette. Places a model nugget containing all the targets in the Models
manager.
Model
Check boxes. Choose which models you want to use in scoring. All the boxes are checked by default. The Check

all and Uncheck all buttons act on all the boxes in a single operation.

Sort by. Enables you to sort the output rows in ascending or descending order for a specified column of the display. The "Selected" option sorts the output based on one or more rows selected by check boxes. This would be useful, for example, to cause target fields named "Market_1" to "Market_9" to be displayed before "Market_10," as
the default sort order displays "Market_10" immediately after "Market_1.&quot؛, View. The default view (Simple) displays the basic set of output columns. The Advanced option displays additional columns for goodness-of-fit measures.

Number of records used in estimation. The number of rows in the original source data file.

Target. The field or fields identified as the target fields (those with a role of Target) in the Type node.

Model. The type of model used for this target field.

Predictors. The number of predictors (those with a role of Input) used for this target
field.

Outliers. This column is displayed only if you have requested (in the Expert Modeler or ARIMA criteria) the automatic detection of outliers. The value shown is the number of outliers detected.

Stationary R-squared. A measure that compares the stationary part of the model to a simple mean model. This measure is preferable to ordinary R-squared when there is a trend or seasonal pattern. Stationary R-squared can be negative with a range of negative infinity to 1. Negative values mean that the model under consideration is worse than the baseline model. Positive values mean that the model under consideration is better than the baseline model.

ISSN: 2537-0367 **eISSN : 2537-0375**

R-Squared. Goodness-of-fit measure of a linear model, sometimes called the coefficient of determination. It is the proportion of variation in the dependent variable explained by the regression model. It ranges in value from 0 to 1. Small values indicate that the model does not fit the data well.

RMSE. Root Mean Square Error. The square root of mean square error. A measure of how much a dependent series varies from its model-predicted level, expressed in the same units as the dependent series.

MAPE. Mean Absolute Percentage Error. A measure of how much a dependent series varies from its model-predicted level. It is independent of the units used and can therefore be used to compare series with different units.

MAE. Mean absolute error. Measures how much the series varies from its model-predicted level. MAE is reported in the original series units.

MaxAPE. Maximum Absolute Percentage Error. The largest forecasted error, expressed as a percentage. This measure is useful for imagining a worst-case scenario for your forecasts.

MaxAE. Maximum Absolute Error. The largest forecasted error, expressed in the same units as the dependent series. Like MaxAPE, it is useful for imagining the worst-case scenario for your forecasts. Maximum absolute error and maximum absolute percentage error may occur at different series points--for example, when the absolute error for a large series value is slightly larger than the absolute error for a small series value. In that case, the maximum absolute error will occur at the larger series value and the maximum absolute percentage error will occur at the smaller series value.

Normalized BIC. Normalized Bayesian Information Criterion. A general measure of the overall fit of a model that attempts to account for model complexity. It is a score based upon the mean square error and includes a penalty for the number of parameters

ISSN: 2537-0367 eISSN : 2537-0375

in the model and the length of the series. The penalty removes the advantage of models with more parameters, making the statistic easy to compare across different models for the same series.

Q. The Ljung-Box Q statistic. A test of the randomness of the residual errors in this model.

df. Degrees of freedom. The number of model parameters that are free to vary when estimating a particular target.

Sig. Significance value of the Ljung-Box statistic. A significance value less than 0.05 indicates that the residual errors are not random.

Summary Statistics. This section contains various summary statistics for the different columns, including mean, minimum, maximum, and percentile values.

The Model Fit table provides fit statistics calculated across all of the models. It provides a concise summary of how well the models, with reestimated parameters, fit the data.

For each statistic, the table provides the mean, standard error (SE), minimum, and maximum value across all models. It also contains percentile values that provide information on the distribution of the statistic across models. For each percentile, that percentage of models have a value of the fit statistic below the stated value. For instance, 95% of the models have a value of MaxAPE (maximum absolute percentage error) that is less than 3.676.

While a number of statistics are reported, we will focus on two: MAPE (mean absolute percentage error) and MaxAPE (maximum absolute percentage error). Absolute percentage error is a measure of how much a dependent series varies from its model- predicted level and provides an indication of the uncertainty in your predictions. The mean absolute percentage error varies from a minimum of 0.669% to a maximum of %١.٠٢٦across all models. The maximum absolute percentage

ISSN: 2537-0367 eISSN : 2537-0375

error varies from 1.742% to 4.373% across all models. So the mean uncertainty in each model's predictions is about 1% and the maximum uncertainty is around 2.5% (the mean value of MaxAPE), with a worst case scenario of about 4%. Whether these values represent an acceptable amount of uncertainty depends on the degree of risk you are willing to accept.

Note : For the descriptive statistics of the model, R-squared represents the coefficient of good fit if the value is greater = 0.96 more than 0.05 this mean the model represent

data exactly (good model).

Model Statistics

Model Number of Predictors

Model Fit statistics Ljung-Box Q(18) Number of

Outliers Stationary R- squared

Statistics DF Sig.

Q. The Ljung-Box Q statistic. A test of the randomness of the residual errors in this model.

df. Degrees of freedom. The number of model parameters that are free to vary when estimating a particular target.

Sig. Significance value of the Ljung-Box statistic. A significance value less than 0.05

indicates that the residual errors are not random.

The Ljung–Box test may be defined as:

H 0 : The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed correlations in the data result from randomness of the sampling process)

H a : The data are not independently distributed; they exhibit serial correlation.

sig more 0.05 accept Null Hypothesis , Data is Random

ISSN: 2537-0367 eISSN : 2537-0375

Results

The Ljung-Box statistic, also known as the modified Box-Pierce statistic, provides an indication of whether the model is correctly specified. A significance value less than implies that there is structure in the observed series which is not accounted for by the model.

The Expert Modeler detected nine points that were considered to be outliers. Each of these points has been modeled appropriately, so there is no need for you to remove them from the series.

Note: (value of Sig. = 0.230) ,by Using residual error test, and when Sig value greater than 0.05 that means the data are random and valid for prediction.

ARIMA Model Parameters

Estimate SE t Sig.

Model_1 الزراعة الغابات وصيد الاسماك No Transformation

Constant 245.208 37.455 6.547 .000

AR

Lag 1 -.963- .048 -19.925- .000

Lag 2 -.918- .062 -14.720- .000

Lag 3 -.949- .042 -22.399- .000

Difference 1

MA Lag 1 -.597- .133 -4.502- .000

The ARIMA model parameters table displays values for all of the parameters in the model, with an entry for each estimated model labeled by the model identifier. For our purposes, it will list all of the variables in the model, including the dependent variable and any independent variables that the Expert Modeler determined were significant. We already know from the model statistics table that there are two significant predictors.

The model parameters table shows us that they are the Number of Catalogs Mailed and the Number of Phone Lines Open for Ordering.

ISSN: 2537-0367 eISSN : 2537-0375

Note : This table provides an estimate of the coefficients of the model, from the model we note that the level of significance Sig= 0.00. Less than 0.05, which indicates that the coefficients are statistically significant, also effective and predictable , the model parameter is significant , able to forecasting

The predicted values show good agreement with the observed values, indicating that the model has satisfactory predictive ability. Notice how well the model predicts the seasonal peaks. And it does a good job of capturing the upward trend of the data.

Note : As in the diagram we observe the compatibility between the observed and real values.

Thus we have predicted a model that represents the data well by using all statistically significant measures , We observed contingency between observed value (red) and real value (blue)

Conclusion

For the descriptive statistics of the model, R-squared represents the coefficient of good fit if the value is greater = 0.96 more than 0.05 this mean the model represent data exactly (good model) This table provides an estimate of the coefficients of the model, from the model we note that the level of significance Sig= 0.00. Less than 0.05, which indicates that the coefficients are statistically significant, also effective and predictable.

ISSN: 2537-0367 eISSN : 2537-0375

References

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. Time series analysis:Forecasting and control, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall.

Brockwell, P. J., and R. A. Davis. 1991. Time Series: Theory and Methods, 2 ed.Springer-Verlag.

Gardner, E. S. 1985. Exponential smoothing: The state of the art. Journal of Forecasting, 4, 1-28.

Harvey, A. C. 1989. Forecasting, structural time series models and the Kalman filter. Cambridge: Cambridge University Press.

Makridakis, S. G., S. C. Wheelwright, and R. J. Hyndman. 1997. Forecasting .Methods and applications, 3rd ed. ed. New York: John Wiley and Sons.

Melard, G. 1984. A fast algorithm for the exact likelihood of autoregressive-moving, average models. Applied Statistics, 33:1, 104-119.

Pena, D., G. C. Tiao, and R. S. Tsay, eds. 2001. A course in time series analysis. New York: John Wiley and Sons. Forecasting: Principles and Practice (2nd ed) Rob J Hyndman and George Athanasopoulos Monash University, Australia
https://otexts.com/fpp2/what-can-be-forecast.html

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on This online. version of the book was last updated on 21 November 2022.The print version of the book (available from Amazon and Google) was last updated on 8 May 2018.

ISSN: 2537-0367 eISSN : 2537-0375

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis (2nd ed). Springer.Introductory Time Series with R (Use R!) 2009th Edition, Kindle Edition https://www.amazon.com/Introductory-Time-R-Use-ebook/dp/B00HWUXJTK/

-https://www.tableau.com/learn/articles/time-series-analysis-books. The Analysis of Time-series: An Introduction with R by Chris Chatfield and Haipeng, Xing Resources in R for Time Series Analysis

-On-line and off-line resources Several Springer Texts in Statistics cover Time Series

Analysis using R, including:

Time Series Analysis & Its Applications: With R Examples (3rd ed)– Robert Shumway and David Stoffer, Springer (2011). In addition Stoffer's own web-site includes a useful R Time Series Tutorial at http://www.stat.pitt.edu/stoffer/tsa2/R_time_series_quick_fix.htm. Introductory Time Series with R Examples – Paul Cowpertwait and Andrew Metcalfe, pringer (2009)

Time Series Analysis in R Examples (2nd ed) – Jonathan Cryer and Kung-Sik Chan, Springer (2009).

Forecasting: Principles & Practice, which uses the forecast package in R. The third, edition, which uses the fable package 2 edition, is also available. https//:otexts.com/fpp2 -/https://data.oecd.org/gdp/domestic-demand-forecast.htm#indicator-chart

ISSN: 2537-0367 eISSN : 2537-0375