Forecasting of the USA Inflation using Different models under Time series

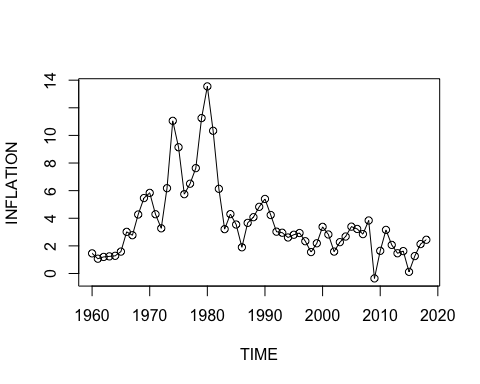
**TERM PAPER**

**Anshuman Patnaik**

**Introduction:**

Understanding and accurately forecasting inflation is crucial for policymakers, economists, and investors. Inflation forecasting enables better decision-making in monetary policies, financial planning, and business strategy. With this study, we aim to investigate USA inflation through time series forecasting models. Specifically, the study explores several models, such as mean forecast, naïve forecast, seasonal naïve, and random walk forecasts. Each model’s accuracy is evaluated, and ARIMA modelling is utilized for more detailed analysis.

**Literature Review:**

Inflation forecasting has been widely studied, with numerous approaches introduced over the years. Traditional models, such as ARIMA (AutoRegressive Integrated Moving Average), have been widely applied due to their capacity to handle non-stationary data and produce reliable short-term forecasts. Advanced models, including seasonal and trend decomposition, are beneficial in capturing seasonal variations often observed in economic time series. Previous research has shown mixed results on which model performs best, with some findings indicating that simple models, such as random walk and naïve forecasts, perform similarly or better than complex ones when dealing with inflation data.

**Research Questions**

1. Which time series model provides the most accurate forecasts for U.S. inflation?
2. How do different forecasting models compare in terms of accuracy and reliability?
3. What are the implications of these forecasts for economic policy?

**Data**

This study analyzes USA inflation data and corresponding GDP values from 1961 to 2018. The dataset was imported and processed in R for model application. Each variable was inspected to ensure it was appropriate for time series analysis.

**Methodology**

Data Transformation and Visualization: The data was loaded and initially visualized to observe patterns. Inflation was converted into a time series object to facilitate time series analysis.

Stationarity Testing: Stationarity, a critical property for time series forecasting, was checked using the Augmented Dickey-Fuller (ADF) test. Since non-stationary data can lead to unreliable models, we differenced the series until it achieved stationarity.

**Modelling Techniques**

Mean Forecast: This approach assumes that future values will match the mean of past observations.

Naïve Forecast: The last observed value is projected forward in this model.

Seasonal Naïve Forecast: Suitable for data with seasonality, this approach repeats previous seasonal patterns.

Random Walk Forecast (RWF): This model assumes that changes in data follow a random path, making it appropriate for non-stationary data with trends.

Evaluation Metrics: Each model was assessed based on key accuracy metrics, including RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error). The Ljung-Box test checked for autocorrelation in residuals, ensuring model robustness.

ARIMA Modeling: ARIMA (AutoRegressive Integrated Moving Average) modeling was applied after testing the initial models. The model was selected using auto. arima() in R, and its effectiveness was tested through residual analysis and forecast comparison.

**Results and Discussion**

The findings indicate that the initial models (mean, naïve, seasonal naïve, and random walk) show varying degrees of accuracy:

Mean Forecast and Naïve Forecast: Both yielded significant errors and autocorrelations, indicating less suitability for this dataset.

Random Walk with Drift: While capturing trends effectively, this model’s residuals indicated possible issues with autocorrelation, limiting its forecasting power.

Seasonal Naïve Forecast: This model exhibited stability with relatively lower error rates, making it suitable for datasets with seasonal variations.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Forecasting Method | Model | Q Stats | Degrees of Freedom (df) | p-value | Significance |
| Mean Model | INFLA1a | 15.93 ***(Worst)*** | 7 | 0.02576 | Significant (p < 0.05) |
| Naïve Method | INFLA1b | 11.281 | 7 | 0.1268 | Insignificant |
| Seasonal Naïve Method | INFLA1c | 11.281 | 7 | 0.1268 | Insignificant |
| RWF | INFLA1d | 11.281 | 7 | 0.1268 | Insignificant |
| ***ARIMA(0,1,2)*** | ***model2*** | ***9.2708*** | ***8 (Best)*** | ***0.32*** | ***Insignificant*** |

**Least RMSE Values for the Predictive models**

**Root Mean Squared Error (RMSE)**: Lower values indicate better prediction accuracy.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Models | INFLA2b | INFLA3c | INFLA4d | ARIMA | INFLA1a |
| RMSE | 1.124240 | 1.124240 | 1.124240 | **1.012214** | **1.012214** |

Both the ARIMA model (model2) and INFLA1a have the lowest RMSE at 1.012214.

While INFLA1a performs well in terms of RMSE and MAE, it exhibits a higher autocorrelation in the residuals. High autocorrelation (observed in the ACF1 metric) in the test set indicates that the errors are not completely random, which suggests the model might be missing some underlying pattern in the data.

This residual autocorrelation can lead to biased forecasts, as it violates the assumption of independently distributed errors. In practical terms, it means that while INFLA1a shows promising accuracy metrics, the presence of autocorrelation suggests it may not fully capture the dynamics of the inflation series, potentially compromising its reliability in longer-term forecasting.

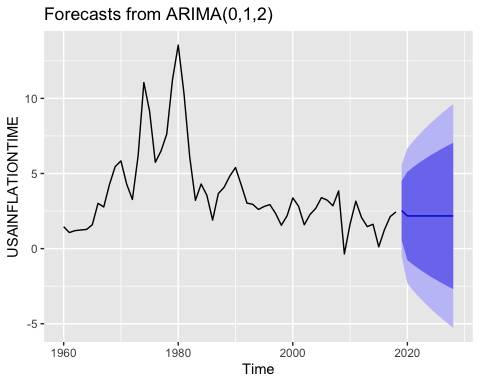
Thus, despite its lower RMSE and MAE, INFLA1a may not be the most robust choice, particularly if capturing independence in residuals is a priority.

***However, testing with the ARIMA model (specifically ARIMA (0,1,2)) demonstrated its efficacy in managing non-stationary data. The model yielded a satisfactory p-value in the Ljung-Box test, confirming that it appropriately handled the dataset’s characteristics.***

***Model - ARIMA(0,1,2)***

**AIC= 219.43** **BIC= 225.61**

|  |  |  |
| --- | --- | --- |
| Coefficients | Estimate | Standard Error |
| MA1 | 0.0956 | 0.1271 |
| MA2 | -0.4019 | 0.1231 |



**Policy Implications**

An accurate inflation forecast becomes important in the formulation of effective economic policies because it bears directly on decisions over the interest rates, monetary policy, and fiscal planning. Reliable forecasts would enable the government to better anticipate inflationary trends in time for interventions that restore stable prices and control inflation. A few key policy implications are listed as follows:

Interest Rate Adjustments: Central banks also depend on inflation forecasts to adjust interest rates, such as what occurs with the Federal Reserve in the United States. If inflation is going to be high, interest rates might be high to limit spending and investment to keep the growth of prices within manageable limits. On the other hand, if low inflation is forecasted to prevail, low interest rates can encourage growth of the economy.

Inflation targeting: Central banks set a target for inflation with the object of keeping stable prices, which acts in turn as the instrument of trust in economic confidence and sustainable growth. Better predictive power would enable policy-makers to make decisions on and sustain those targets better, thereby upholding greater public confidence in monetary policy and reducing uncertainty among investors, businesses, and consumers.

Resource Allocation: Inflation forecasts will allow government policymakers to allocate more effective budgets within likely frameworks. For example, if it is predicted that inflationary rates will rise severely in certain areas of the economy, such as food or energy, then policymakers will automatically offer more subsidized policies or relief measures to soften the impact for low-income families.

Public Expectations Management: The correct inflation forecasts and the efficient communication of such expectations act as an instrument of influence over the public perception and conduct, which could then be transmitted to actual inflation via mechanisms like wage negotiations and spending by consumers.

Crisis Prevention: Good inflation forecasting will provide insight into hyperinflation or deflation, allowing actions to be undertaken before a situation culminates into a full-fledged economic crisis. Such insight will be particularly invaluable in times of crisis-like economic shock as in financial crises or a disruption in supply chains.

**Future Research**

Although this paper has demonstrated the applicability of several time series models to inflation forecasting, areas exist which would further enhance the accuracy and applicability of the forecast:

Further Extension of Macroeconomic Variables: Other macroeconomic variables such as GDP, unemployment rate, exchange rate and commodity prices, etc., provide scope for further extending the application of the model. These variables are often related to inflation. Including these variables in a VAR or SEMs model might enhance the power of the models.

Advanced versions of machine learning models such as neural networks, random forests, and support vector machines have a lot of promise for being applied in time series forecasting. Future studies can compare machine learning models with traditional time series models to see if they deliver superior performance in inflationary forecasts.

Real-time data analysis. Of course, there are the usual caveats: such estimates depend on rather old data. Future studies can examine integration with real-time data sources-for example, via web scraping to feed inflation information or satellite imagery as inputs for some supply chains. This would make inflation forecasts more responsive and timely.

Inclusion of High Frequency Data: Many of the economic indicators are reported quarterly or monthly. However, due to increased gains in data availability, high frequency data (daily or weekly) could be included too. The usage of high-frequency data improves the accuracy of short term forecast in inflation; it is able to provide policymakers with just-in-time insights.

Structural Break Accounting: Inflation dynamics can be structural in terms of monetary policy shifts, bilateral and multilateral international trade agreements, and changes in technology. Methods for the identification and accounting of such structural breaks can be developed for inflation forecasting models, which would improve model adaptability and accuracy over time.

Regional Inflation Forecasting: Regional regional regional regional. Regional regional regional regional regional differences in inflation between regions are most pronounced because economic conditions, policies, and demographics vary across regions within a country. Further research can be explored concerning regional inflation forecasting with such much more localized insights for the design of targeted economic policies.

**Conclusion**

The paper argues that time series models can be useful for inflation forecasting and the determination of monetary and economic stability policies. However, with advancing improvements in data availability and algorithms, future study can base ongoing forecasting methodologies on the enhancement of accuracy, as well as applicability, of inflation forecasts. These efforts will further aid policymakers in making informed decisions, thus facilitating sustainable economic growth as well as price stability.

**R CODES**

options(repos = c(CRAN = "https://cran.rstudio.com/"))

#INSTALLING PACKAGES   
library(ggplot2)  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ lubridate 1.9.3 ✔ tibble 3.2.1  
## ✔ purrr 1.0.2 ✔ tidyr 1.3.1  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

install.packages("forecast")

##   
## The downloaded binary packages are in  
## /var/folders/w4/mp0tg4c9103\_lzs8q55f06h40000gn/T//RtmpvdaKHK/downloaded\_packages

library(forecast)

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

install.packages("tseries")

##   
## The downloaded binary packages are in  
## /var/folders/w4/mp0tg4c9103\_lzs8q55f06h40000gn/T//RtmpvdaKHK/downloaded\_packages

library(tseries)  
library(vars)

## Loading required package: MASS  
##   
## Attaching package: 'MASS'  
##   
## The following object is masked from 'package:dplyr':  
##   
## select  
##   
## Loading required package: strucchange  
## Loading required package: zoo  
##   
## Attaching package: 'zoo'  
##   
## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric  
##   
## Loading required package: sandwich  
##   
## Attaching package: 'strucchange'  
##   
## The following object is masked from 'package:stringr':  
##   
## boundary  
##   
## Loading required package: urca  
## Loading required package: lmtest

install.packages("fpp2")

##   
## The downloaded binary packages are in  
## /var/folders/w4/mp0tg4c9103\_lzs8q55f06h40000gn/T//RtmpvdaKHK/downloaded\_packages

library(fpp2)

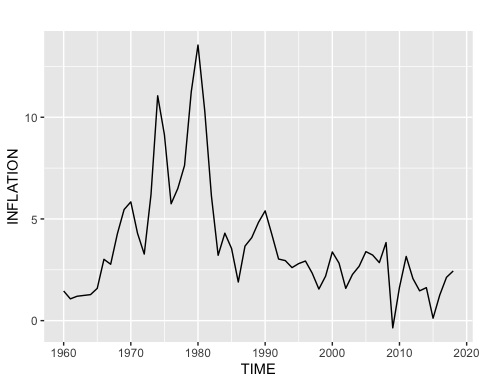
## ── Attaching packages ────────────────────────────────────────────── fpp2 2.5 ──  
## ✔ fma 2.5 ✔ expsmooth 2.3

install.packages("readxl")

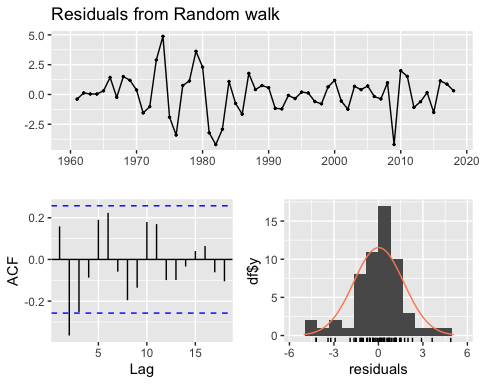
##   
## The downloaded binary packages are in  
## /var/folders/w4/mp0tg4c9103\_lzs8q55f06h40000gn/T//RtmpvdaKHK/downloaded\_packages

library(readxl)  
USA\_GDP\_WORKING\_ <- read\_excel("/Users/anshuman/Desktop/TERM PAPER, BAIJU/DATASET/USA/USAGDPWORKING1.xls")  
  
  
#VIEWING THE DATASET USA\_GDP\_WORKING (includes inflation)  
  
NewDATA=USA\_GDP\_WORKING\_[1:59,c(1,3)]  
  
  
#PLOTTING THE VARIABLES OF THE DATASET  
plot(NewDATA$TIME,y=NewDATA$INFLATION,type = ("o"),xlab=("TIME"),ylab=("INFLATION"))

#converting inflation into timeseries dataset  
  
USAINFLATIONTIME=ts(NewDATA$INFLATION,start = min(NewDATA$TIME),end = max(NewDATA$TIME),frequency = 1)  
  
#time plots of the datasets  
autoplot(USAINFLATIONTIME)+xlab("TIME")+ylab("INFLATION")

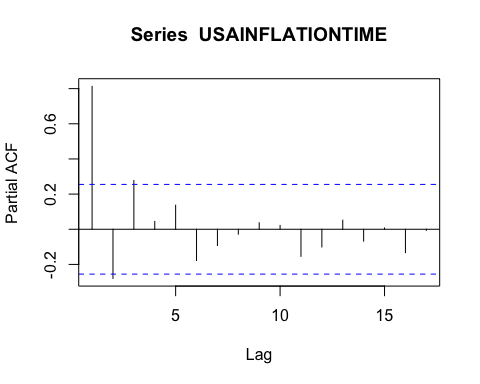


#CHECKING RESIDUALS FOR OUR INFLATION  
checkresiduals(rwf(USAINFLATIONTIME))



##   
## Ljung-Box test  
##   
## data: Residuals from Random walk  
## Q\* = 26.575, df = 10, p-value = 0.003039 **(HIGH AUTO CORR)**  
##   
## Model df: 0. Total lags used: 10

#autocorrelation check for the data  
  
pacf(USAINFLATIONTIME)



#checking stationary USAINFLATIONTIME  
adf.test(USAINFLATIONTIME)

##   
## Augmented Dickey-Fuller Test  
##   
## data: USAINFLATIONTIME  
## Dickey-Fuller = -2.5748, Lag order = 3, p-value = 0.3425 **(NO STATIONARY)**  
## alternative hypothesis: stationary

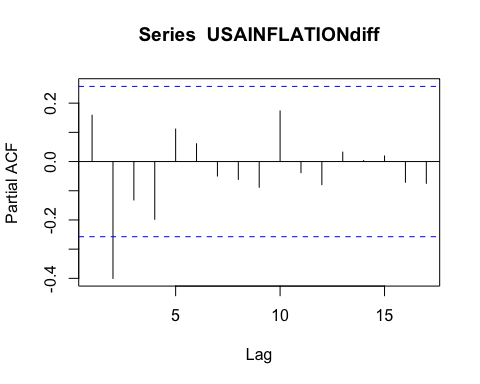
USAINFLATIONdiff=diff(USAINFLATIONTIME)

#checking stationary for USAINFLATIONdiff  
adf.test(USAINFLATIONdiff)

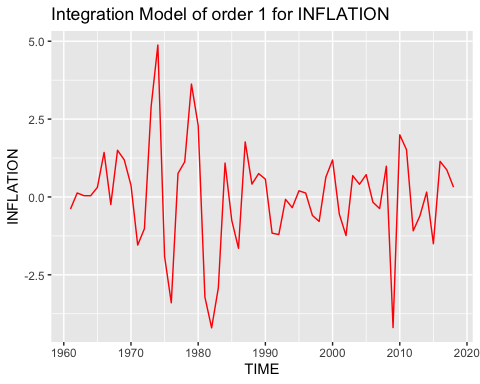
## Warning in adf.test(USAINFLATIONdiff): p-value smaller than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: USAINFLATIONdiff  
## Dickey-Fuller = -5.5041, Lag order = 3, p-value = 0.01 **(STATIONARY)**  
## alternative hypothesis: stationary

#checking if white noise is occurring   
pacf(USAINFLATIONdiff)



#AUTOPLOT FOR THE DIFFERENCED VARIABALES  
autoplot(USAINFLATIONdiff,xlab=("TIME"),ylab = ("INFLATION"),col=c("red"))+ggtitle("Integration Model of order 1 for INFLATION")



#CREATING A TRAINING DATA FOR CHECKING THE ACCURACY OF OUR FORECASTING MODELS   
#for inflation data  
USAINFLATIONTIME\_2=window(USAINFLATIONdiff,start=1961,end=1995)  
INFLA1=meanf(USAINFLATIONTIME\_2,h=10)  
INFLA2=naive(USAINFLATIONTIME\_2,h=10)  
INFLA3=snaive(USAINFLATIONTIME\_2,h=10)  
INFLA4=rwf(USAINFLATIONTIME\_2,h=10,drift = FALSE)  
  
  
  
#creating a test data and checking accuracy  
USAINFLATIONTIME\_3=window(USAINFLATIONdiff,start=1996)  
  
accuracy(INFLA1,USAINFLATIONTIME\_3)

## ME RMSE MAE MPE MAPE MASE  
## Training set -6.344132e-18 1.8925713 1.4064432 94.29607 94.29607 0.8690652  
## Test set 2.023432e-02 0.7588745 0.6849832 96.11188 96.11188 0.4232628  
## ACF1 Theil's U  
## Training set 0.29458906 NA  
## Test set 0.07110285 0.9567035

accuracy(INFLA2,USAINFLATIONTIME\_3)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.01721265 2.2794443 1.618340 55.66419 198.91278 1.0000000  
## Test set -0.13924538 0.7712784 0.667526 80.00531 91.48424 0.4124757  
## ACF1 Theil's U  
## Training set -0.01218417 NA  
## Test set 0.07110285 1.113384

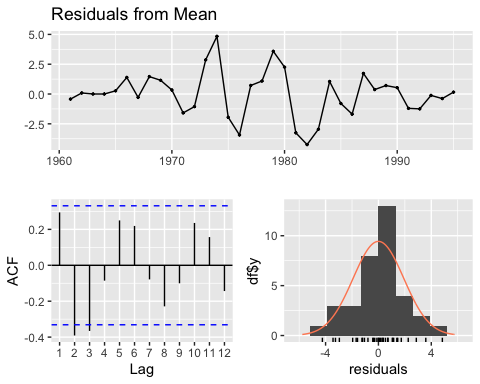
accuracy(INFLA3,USAINFLATIONTIME\_3)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.01721265 2.2794443 1.618340 55.66419 198.91278 1.0000000  
## Test set -0.13924538 0.7712784 0.667526 80.00531 91.48424 0.4124757  
## ACF1 Theil's U  
## Training set -0.01218417 NA  
## Test set 0.07110285 1.113384

accuracy(INFLA4,USAINFLATIONTIME\_3)

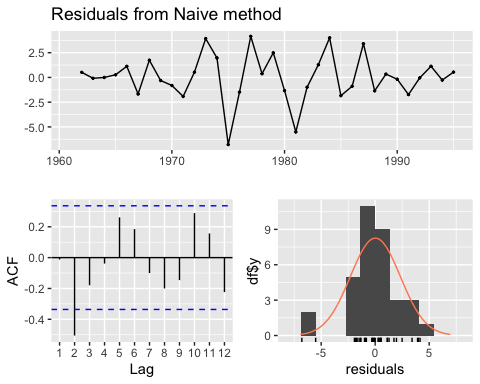
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.01721265 2.2794443 1.618340 55.66419 198.91278 1.0000000  
## Test set -0.13924538 0.7712784 0.667526 80.00531 91.48424 0.4124757  
## ACF1 Theil's U  
## Training set -0.01218417 NA  
## Test set 0.07110285 1.113384

#checking for autocorrealtions  
checkresiduals(INFLA1) **#significant p value**



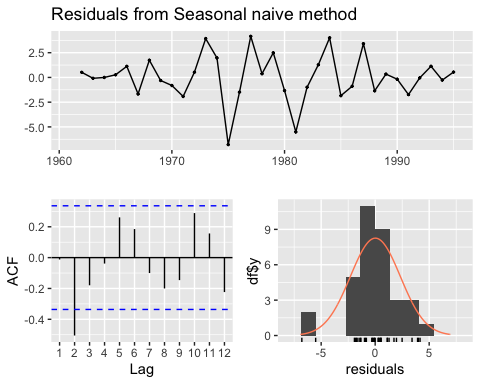
##   
## Ljung-Box test  
##   
## data: Residuals from Mean  
## Q\* = 20.149, df = 7, p-value = 0.005257  
##   
## Model df: 0. Total lags used: 7

checkresiduals(INFLA2) **#significant p value**



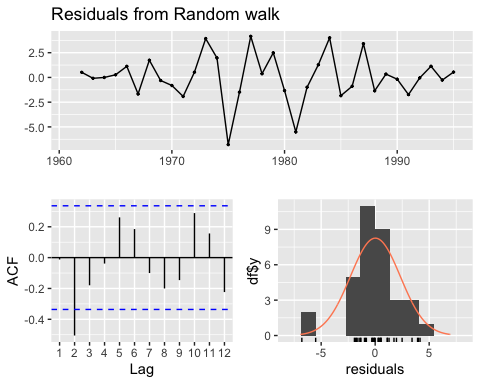
##   
## Ljung-Box test  
##   
## data: Residuals from Naive method  
## Q\* = 15.93, df = 7, p-value = 0.02576  
##   
## Model df: 0. Total lags used: 7

checkresiduals(INFLA3) **#significant p value**



##   
## Ljung-Box test  
##   
## data: Residuals from Seasonal naive method  
## Q\* = 15.93, df = 7, p-value = 0.02576  
##   
## Model df: 0. Total lags used: 7

checkresiduals(INFLA4) **#significant p value**



##   
## Ljung-Box test  
##   
## data: Residuals from Random walk  
## Q\* = 15.93, df = 7, p-value = 0.02576  
##   
## Model df: 0. Total lags used: 7

**DUE TO THE SIGNIFICANT P VALUE< WE REJECT THE NULL OF NO AUTO CORRELATION IN THE FORECAST MODELS. WE THUS CREATE A NEW DATA SET WITH SECOND DIFFERENCE OF THE INFLATION DATA TO FORECAST OUR MODELS**

#checking for Second Difference   
#CREATING A TRAINING DATA FOR CHECKING THE ACCURACY OF OUR FORECASTING MODELS   
#for inflation data  
USAINFLATIONTIME\_2=window(diff(USAINFLATIONdiff),start=1961,end=1995)

## Warning in window.default(x, ...): 'start' value not changed

INFLA1a=meanf(USAINFLATIONTIME\_2,h=10)  
INFLA2b=naive(USAINFLATIONTIME\_2,h=10)  
INFLA3c=snaive(USAINFLATIONTIME\_2,h=10)  
INFLA4d=rwf(USAINFLATIONTIME\_2,h=10,drift = FALSE)  
  
  
#creating a test data and checking accuracy  
USAINFLATIONTIME\_3=window(diff(USAINFLATIONdiff),start=1996)  
  
accuracy(INFLA1a,USAINFLATIONTIME\_3)

## ME RMSE MAE MPE MAPE MASE  
## Training set 7.510332e-17 2.279379 1.6193528 132.3893 132.3893 0.6120015  
## Test set 3.454056e-02 1.012214 0.7930518 103.4107 103.4107 0.2997178  
## ACF1 Theil's U  
## Training set -0.01218417 NA  
## Test set -0.09983095 0.9658152

accuracy(INFLA2b,USAINFLATIONTIME\_3)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.0008149222 3.289468 2.6459948 -155.8666 518.2452 1.0000000  
## Test set -0.4904402650 1.124240 0.9448129 207.4366 222.6019 0.3570729  
## ACF1 Theil's U  
## Training set -0.25478528 NA  
## Test set -0.09983095 1.358236

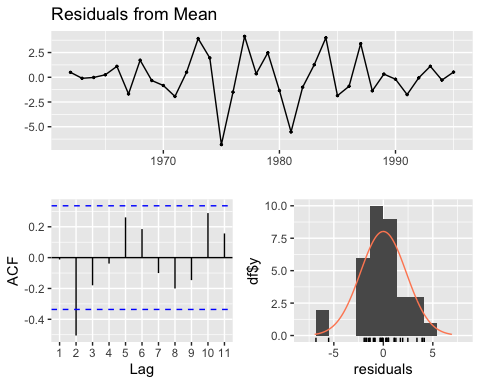
accuracy(INFLA3c,USAINFLATIONTIME\_3)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.0008149222 3.289468 2.6459948 -155.8666 518.2452 1.0000000  
## Test set -0.4904402650 1.124240 0.9448129 207.4366 222.6019 0.3570729  
## ACF1 Theil's U  
## Training set -0.25478528 NA  
## Test set -0.09983095 1.358236

accuracy(INFLA4d,USAINFLATIONTIME\_3)

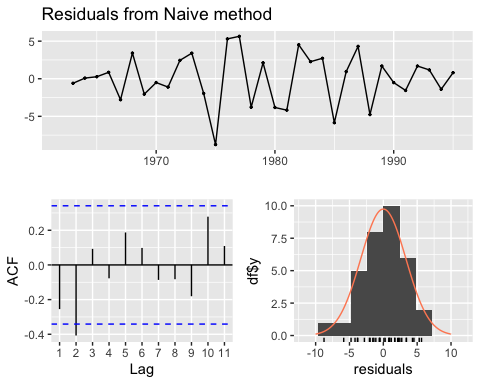
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.0008149222 3.289468 2.6459948 -155.8666 518.2452 1.0000000  
## Test set -0.4904402650 1.124240 0.9448129 207.4366 222.6019 0.3570729  
## ACF1 Theil's U  
## Training set -0.25478528 NA  
## Test set -0.09983095 1.358236

#checking for autocorrealtions  
checkresiduals(INFLA1a)#significant p value



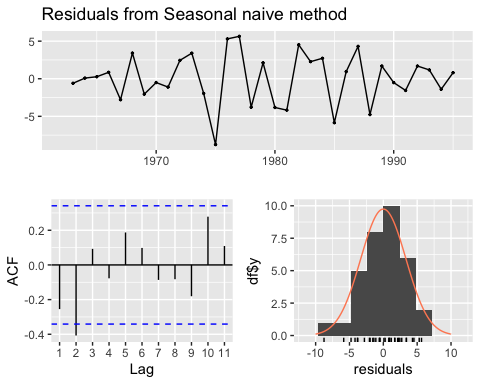
##   
## Ljung-Box test  
##   
## data: Residuals from Mean  
## Q\* = 15.93, df = 7, p-value = 0.02576  
##   
## Model df: 0. Total lags used: 7

checkresiduals(INFLA2b)#insignificant p value



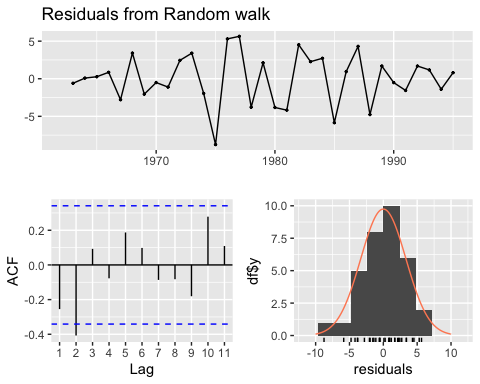
##   
## Ljung-Box test  
##   
## data: Residuals from Naive method  
## Q\* = 11.281, df = 7, p-value = 0.1268  
##   
## Model df: 0. Total lags used: 7

checkresiduals(INFLA3c)#insignificant p value



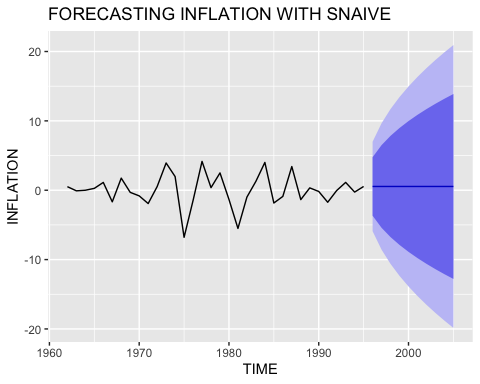
##   
## Ljung-Box test  
##   
## data: Residuals from Seasonal naive method  
## Q\* = 11.281, df = 7, p-value = 0.1268  
##   
## Model df: 0. Total lags used: 7

checkresiduals(INFLA4d)#insignificant p value



##   
## Ljung-Box test  
##   
## data: Residuals from Random walk  
## Q\* = 11.281, df = 7, p-value = 0.1268  
##   
## Model df: 0. Total lags used: 7

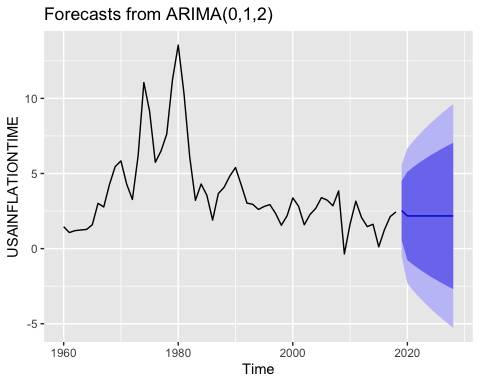
#out of this, INFLA3 is the most stable form of prediction  
autoplot(INFLA4d)+xlab("TIME")+ylab("INFLATION")+ggtitle("FORECASTING INFLATION WITH SNAIVE")



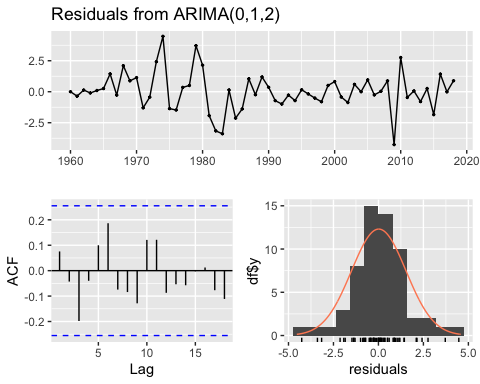
#Through the ARIMA MODELLING   
  
#INFLATION FORECASTING (DIRECTLY WITHOUT STATIONARITY)  
  
model2=auto.arima(USAINFLATIONTIME)  
model2

## Series: USAINFLATIONTIME   
## ARIMA(0,1,2)   
##   
## Coefficients:  
## ma1 ma2  
## 0.0956 -0.4019  
## s.e. 0.1271 0.1231  
##   
## sigma^2 = 2.388: log likelihood = -106.71  
## AIC=219.43 AICc=219.87 BIC=225.61

forecast2=forecast(model2,h=10)  
autoplot(forecast2)



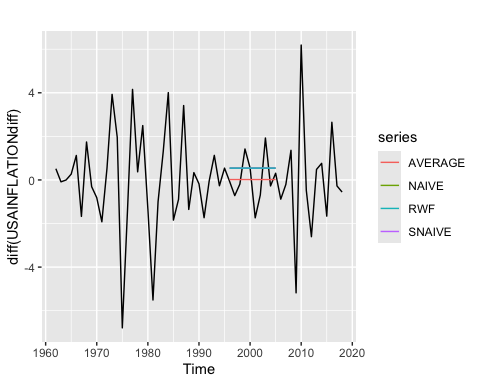
checkresiduals(model2) #Insignificant P value (model is correct)

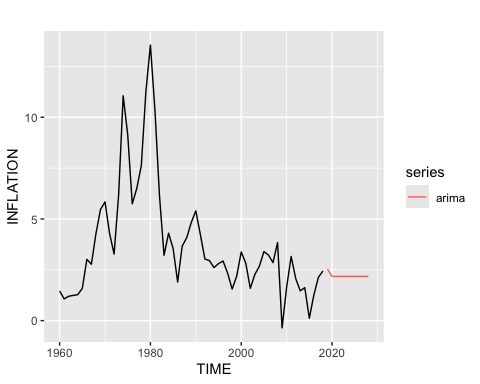


##   
## Ljung-Box test  
##   
## data: Residuals from ARIMA(0,1,2)  
## Q\* = 9.2708, df = 8, p-value = 0.32  
##   
## Model df: 2. Total lags used: 10

#comparing forecasts  
  
autoplot(diff(USAINFLATIONdiff))+  
 autolayer(INFLA1a$mean,series="AVERAGE")+  
 autolayer(INFLA2b$mean,series="NAIVE")+  
 autolayer(INFLA3c$mean,series="SNAIVE")+  
 autolayer(INFLA4d$mean,series="RWF")

autoplot(USAINFLATIONTIME)+ autolayer(forecast2$mean,series="arima")+xlab("TIME")+  
 ylab("INFLATION")





**References**

Maddala, G. S. (1992). *Introduction to Econometrics* (2nd ed.). Macmillan.

Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach* (7th ed.). Cengage Learning.