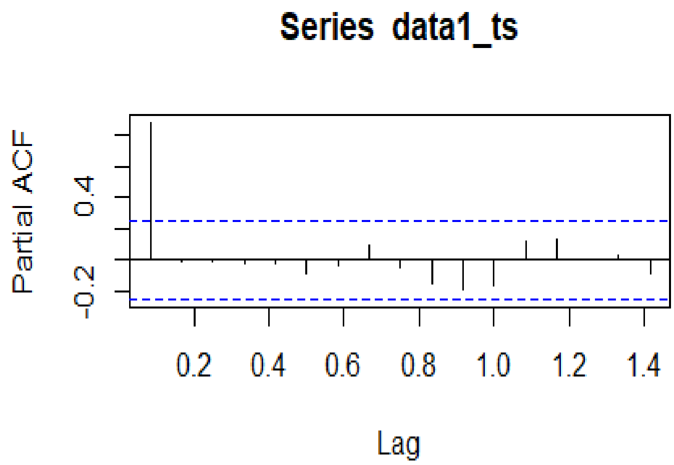
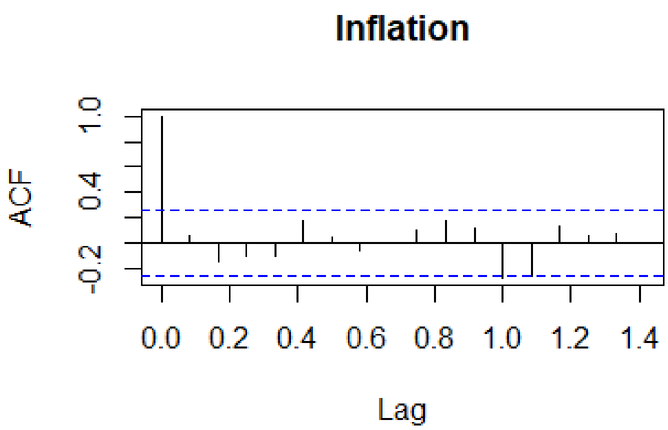
Trimester: Oct – Dec 24

Maximum Marks: 50 Examination: ETE Exam Date: 17/01/2025 Duration: 3 Hours		
Programme code: 01 **Programme: MBA (BADS Minor)**	**Class:** SY	**Semester/Trimester: V**
College: K. J. Somaiya Institute of Management	**Name of the department/Section/Center: Business Analytics**	
Course Code: 217P01M532	**Name of the Course: Predictive Analytics**	

Instructions:

1. **All questions are compulsory.**
2. **Make suitable assumptions if required and state them.**

Question No.		**Max. Marks**
1	Here are the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs for the inflation data:   a) Based on the ACF plot, do you observe any significant lags with high autocorrelation? How would this influence your model	10

choice?

b) Looking at the PACF plot, do you see a sharp cutoff after a specific lag? What does that suggest about the AR process of the data?

c) If the ACF shows a slow decay, what type of model would you consider for this time series? Would you prefer an AR model or an MA model, and why?

d) In the context of time series modelling, how do you use the ACF and PACF plots to determine the parameters for ARIMA or similar models?

2 — Following is the R code for the SARIMA model. — 20

```
setwd("D:/Predictive Modelling/triple exponential smoothing")
y<-read.csv("sales goods company.csv")
yt<-ts(y, start=2012, frequency=4)
plot(yt, col="black")
trend<-lm(yt ~ time(yt))
abline(trend, col="red")

adf.test(yt)
ytd1<-diff(yt , k=1)
adf.test(ytd1)
acf(ytd1) # q=0
pacf(ytd1) # p=1
arima(yt, order=c(1,1,0))
arima(yt, order=c(2,1,0))
arima(yt, order=c(1,2,1))

yt_seasonal<-diff(yt, k=4) #
adf.test(yt_seasonal)  # D=0
acf(yt_seasonal)  # Q=0
pacf(yt_seasonal)  # P=2

arima(yt, order=c(2,2,1),  seasonal = list(order=c(2,0,0), period=4))

model1<-arima(yt, order=c(2,2,1),  seasonal = list(order=c(2,1,0), period=4))
arima(yt, order=c(2,2,1),  seasonal = list(order=c(2,1,1), period=4))

library("forecast")
auto.arima(yt)


Box.test(residuals(model1), lag=10, type="Ljung-Box")
```

Give the answer to the following questions based on the above code

a) What is the purpose of using ts() function in the code? How does it transform the data?

b) What does the Augmented Dickey-Fuller test (adf.test()) assess in time series analysis? Frame the Null and Alternative hypotheses for this test. If the p-value for this test is 0.257, what conclusion would you draw, and what would be the next step to perform the analysis?

c) How does the ARIMA model (1,1,0) differ from (2,1,0) and (1,2,1) in terms of model complexity and interpretation?

d) What do the parameters order=c(p,d,q) represent in the ARIMA model?

e) How does the Box.test(residuals(model1), lag=10, type="Ljung-Box") function help in evaluating the model's residuals? What would indicate a good model fit? Frame the hypotheses.

3 — Following is the R code for exponential smoothing. — 20

```
# Set working directory and load the dataset
```

```
setwd("D:/time series forecasting/ expo smoothing")

# Load data
demand_data <- read.csv("weekly_demand.csv")

# Convert to time series
demand_ts <- ts(demand_data, start = c(2020, 1), frequency = 52)

# Plot the original time series
plot(demand_ts, main = "Weekly Demand Time Series", col = "blue", pch = 19)

# Load the forecast package
install.packages("forecast")
library(forecast)

# Apply exponential smoothing
ets_model <- ets(demand_ts, model="ANN")
print("Exponential Smoothing Model:")
print(ets_model)

# Plot the forecast
plot(ets_model, main = " Exponential Smoothing Forecast", col = "darkgreen")

# Calculate residuals and MSE
residuals_ets <- residuals(ets_model)
mse_ets <- mean(residuals_ets^2)
print(paste("MSE for  Exponential Smoothing:", mse_ets)
```

Give the answer to the following questions based on the above code

a) What steps would you take if there are missing values in the demand_data before converting it into a time series?
b) Explain the meaning of the "ANN" model components in the ets() function.
c) In what scenarios might the "ANN" model be insufficient for capturing the patterns in the data?
d) What is the significance of residuals in evaluating the performance of the ets_model?
e) If you wanted to implement exponential smoothing for data with daily frequency, what changes would you make to the code?