**HW-4**

CODE:

setwd("F:/LUC Study Material/ISSCM 495")

getwd()

hpi=read.csv("hpi.csv",sep="|")

hpi.ss=hpi[c("YEAR","MA.Boston")]

install.packages("stringr")

library(stringr)

hpi.ss$YEAR=str\_replace(str\_replace(hpi.ss$YEAR," ",""),"^","01")

hpi.ss$YEAR=as.Date(hpi.ss$YEAR,"%d%B%Y")

plot(hpi.ss,type='l')

DataframeTS = function(x,y) {

styr=as.numeric(format((head(x[1],1)),"%Y"))

stmth=as.numeric(format((head(x[1],1)),"%m"))

endyr=as.numeric(format((tail(x[1],1)),"%Y"))

endmth=as.numeric(format((tail(x[1],1)),"%m"))

return(ts(x[2],start=c(styr,stmth),end=c(endyr,endmth),frequency=y))

}

hpi.ts=DataframeTS(hpi.ss,12)

plot.ts(hpi.ts,type='l')

tsScale=function(x){

fe=x[1]

return((x/fe)\*100)

}

hpi.win=window(hpi.ts,start=2008,end=2015)

hpi.win=tsScale(hpi.win)

plot(hpi.win)

hpi.win=window(hpi.ts,start=2010+.25,end=2015)

hpi.win=tsScale(hpi.win)

lag.plot(hpi.win,lag=9,do.lines=FALSE)

install.packages("forecast")

library("forecast")

require(tseries)

acf(hpi.win)

seasonplot(hpi.win,year.labels=TRUE, year.labels.left = TRUE,col=1:3)

monthplot(hpi.win)

hpi.test=window(hpi.win,start=2014+.5)

hpi.train=window(hpi.win,end=2014+.45)

############################################################################################################################

## idea of differencing

############################################################################################################################

hpi.diff=diff(hpi.ts,1,2)

plot(hpi.diff)

hpi.diff=diff(hpi.ts,12,1)

plot(hpi.diff)

hpi.diff=diff(diff(BoxCox(hpi.ts,lambda = BoxCox.lambda(hpi.ts)),12,1),1,2)

plot(hpi.diff)

adf.test(hpi.diff,alternative="stationary")

kpss.test(hpi.diff)

hpi.diff=diff(hpi.train,1,1)

kpss.test(hpi.diff)

adf.test(hpi.diff)

Acf(hpi.diff,main="acf",lag.max=72)

Pacf(hpi.diff,main="pacf")

fit=Arima(hpi.train, order=c(2,1,1))

fit

res=residuals(fit)

hist(res)

Box.test(res,lag=10,fitdf=3,type="L")

##Forecast for 12 months##

fc=forecast(fit,h=12)

plot(fc)

lines(hpi.test,col="green")

accuracy(fc,hpi.test)

##Simpler method: auto estimate p, d and q##

fit\_auto=auto.arima(hpi.train)

## lets see the model.

fit\_auto

#### it gives arima(2,1,0) as ideal model but with just in one step.

fc\_auto=forecast(fit\_auto,h=12)

plot(fc\_auto)

lines(hpi.test,col="green")

accuracy(fc\_auto,hpi.test)

############################################################################################################################

##Lets take another example of airlines from mid-term.

############################################################################################################################

air=read.csv("a.csv",sep=",",header=TRUE)

air$Month=(str\_replace(air$Month,"^","01-"))

air$Month=as.Date(air$Month, "%d-%Y-%m")

plot(air, type='l')

DataframeTS = function(x,y) {

styr=as.numeric(format((head(x[1],1)),"%Y"))

stmth=as.numeric(format((head(x[1],1)),"%m"))

endyr=as.numeric(format((tail(x[1],1)),"%Y"))

endmth=as.numeric(format((tail(x[1],1)),"%m"))

return(ts(x[2],start=c(styr,stmth),end=c(endyr,endmth),frequency=y))

}

air.ts=DataframeTS(air,12)

lag.plot(air.ts,lag=9,do.lines=FALSE)

library("forecast")

acf(air.ts)

seasonplot(air.ts,year.labels=TRUE, year.labels.left = TRUE,col=1:12)

monthplot(air.ts)

air.test=window(air.ts,start=1960)

air.train=window(air.ts,end=1960-.001)

air.train

air.test

air.diff=diff(air.train,1,1)

acf(air.diff)

air.diff=diff(air.diff,12,1)

plot(air.diff)

acf(air.diff)

## lets check adf test to confirm that we have achieved stationarity.

kpss.test(air.diff)

adf.test(air.diff)

tsdisplay(air.diff)

fit=Arima(air.train, order=c(1,1,1), seasonal=c(0,1,0))

fit

res=residuals(fit)

Box.test(res, lag=24, fitdf=2,type="L")

## large p value, residuals are random.

hist(res)

hist(res,breaks=20)

## forecasting.

fc=forecast(fit,h=12)

plot(fc)

lines(air.test,col="red")

accuracy(fc,air.test)

## auto.arima function.

fit\_auto=auto.arima(air.train)

fit\_auto

fc\_auto=forecast(fit\_auto,h=12)

plot(fc\_auto)

lines(air.test,col="red")

accuracy(fc\_auto,air.test)

############################################################################################################################

##Lets take another example of female unemployment from mid-term.

############################################################################################################################

uemp=read.csv("fn.csv",sep="|",header=TRUE)

summary(uemp)

##1948-01 is the format, So we need to add 01 at the end. $ denotes the end of line in regex.

uemp$Month=str\_replace(uemp$Month,"$","-01")

## convert the string to date.

uemp$Month=as.Date(uemp$Month, "%Y-%m-%d")

## plot the data.

plot(uemp,type='l')

uemp.ts=DataframeTS(uemp,12)

uemp.ts=tsScale(uemp.ts)

## training and test datasets

uemp.train=window(uemp.ts,start=1975,end=1981-.001)

uemp.test=window(uemp.ts,start=1981)

## see how test and training datasets look.

uemp.train

uemp.test

adf.test(uemp.train)

uemp.diff=diff(uemp.train)

adf.test(uemp.diff)

kpss.test(uemp.diff)

Acf(uemp.diff)

Pacf(uemp.diff)

Acf(uemp.diff,lag.max=72)

Pacf(uemp.diff,lag.max=72)

uemp.diff=diff(uemp.diff,12,1)

## Now see the acf.

Acf(uemp.diff,lag.max=72)

pacf(uemp.diff,lag.max=72)

Acf(uemp.diff,lag.max=5)

Pacf(uemp.diff,lag.max=5)

fit=Arima(uemp.train, order=c(1,1,2),seasonal=c(1,1,1))

res=residuals(fit)

plot(res)

hist(res)

## Check box ljung test.

Box.test(res, lag=24, fitdf=6,type="L")

fc=forecast(fit,h=12)

plot(fc)

lines(uemp.test)

accuracy(fc,uemp.test)

fit=arima(uemp.train, order=c(1,1,0),seasonal=c(1,1,0))

## check the residuals.

res=residuals(fit)

plot(res)

hist(res)

## Check box ljung test.

Box.test(res, lag=12, fitdf=6,type="L")

## P value is 40% so data is random.

fc=forecast(fit,h=12)

plot(fc)

lines(uemp.test)

accuracy(fc,uemp.test)

fit=arima(uemp.train, order=c(1,1,0),seasonal=c(2,1,0))

## check the residuals.

res=residuals(fit)

plot(res)

hist(res)

## Check box ljung test.

Box.test(res, lag=12, fitdf=5,type="L")

## P value is 40% so data is random.

fc=forecast(fit,h=12)

plot(fc)

lines(uemp.test)

accuracy(fc,uemp.test)

fit=arima(uemp.train, order=c(1,1,0),seasonal=c(2,0,0))

## check the residuals.

res=residuals(fit)

plot(res)

hist(res)

## Check box ljung test.

Box.test(res, lag=12, fitdf=4,type="L")

fc=forecast(fit,h=12)

plot(fc)

lines(uemp.test)

accuracy(fc,uemp.test)

fit=auto.arima(uemp.train)

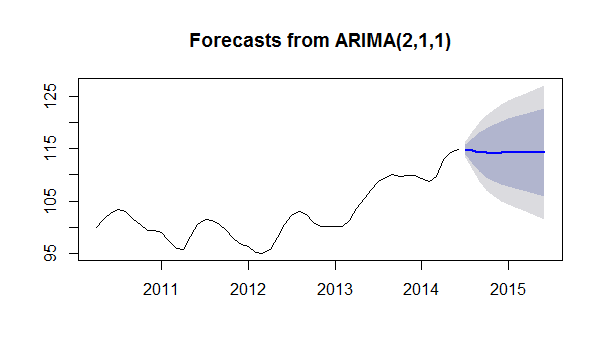
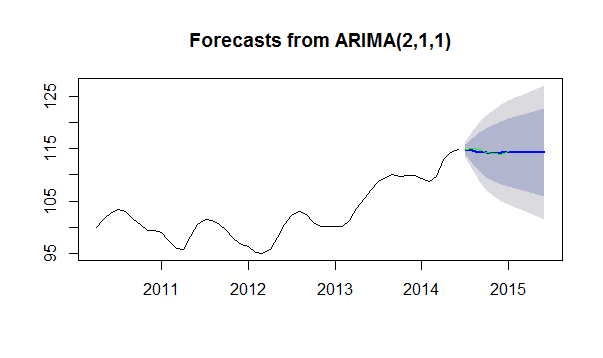
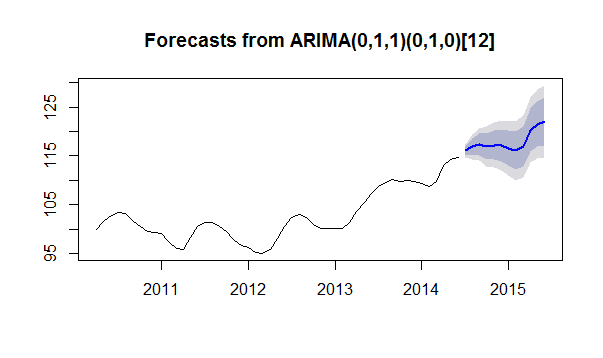
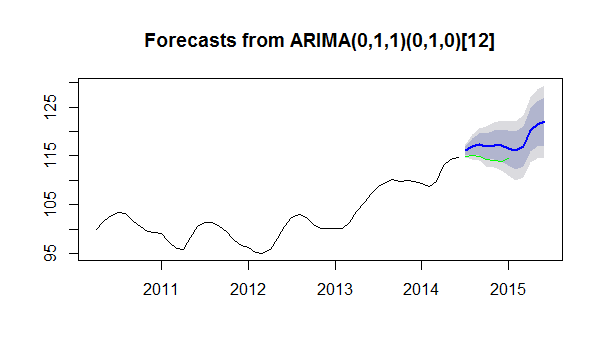
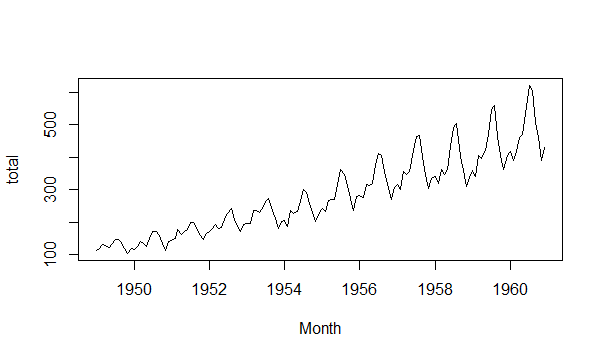
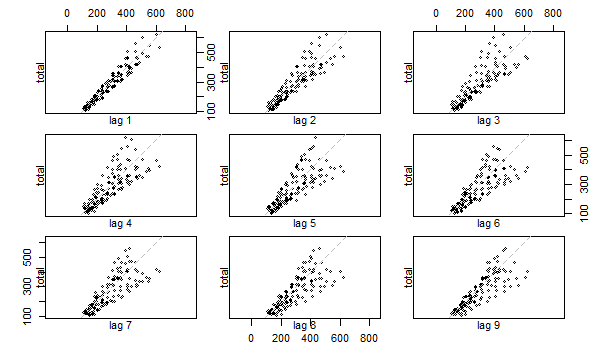
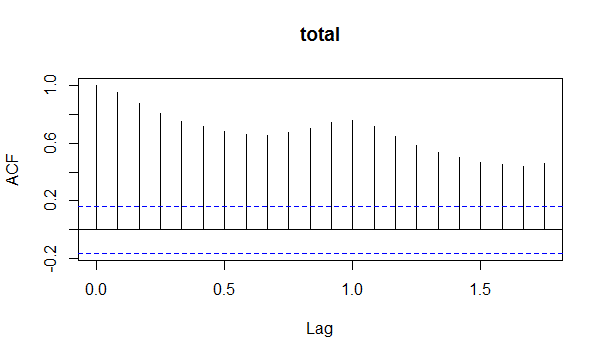
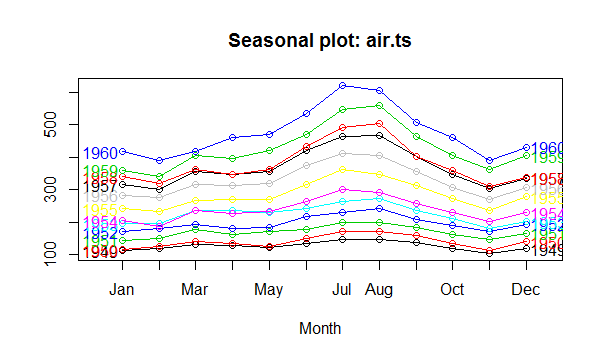
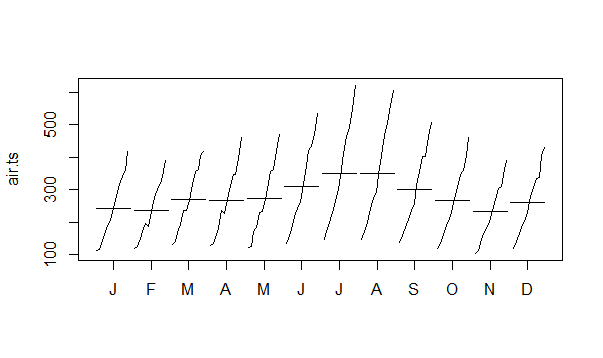
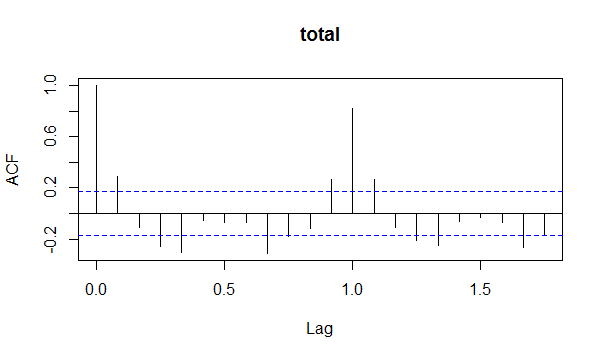
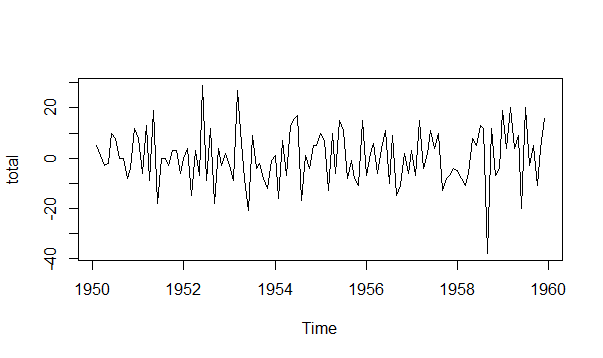
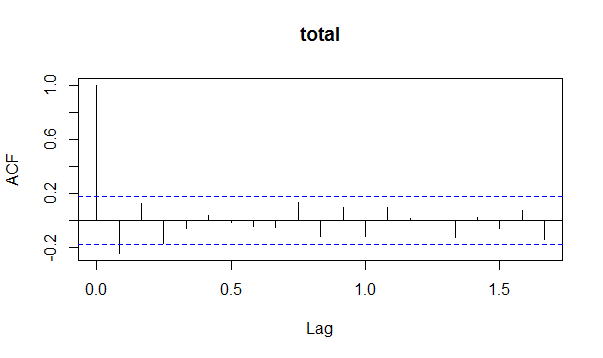
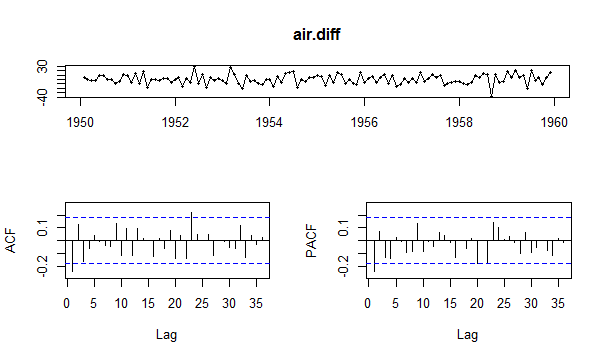
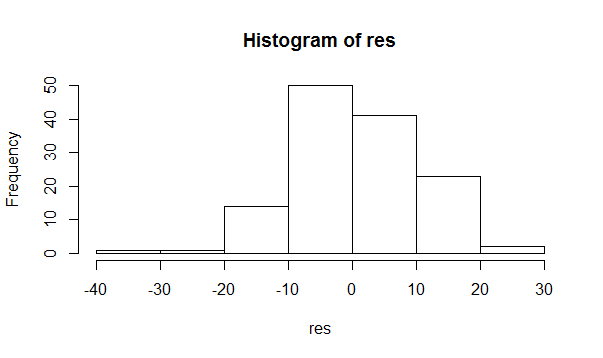
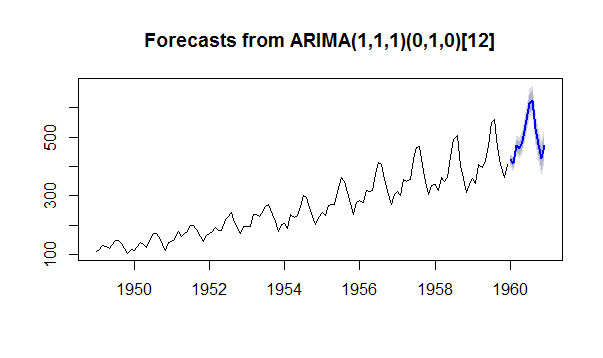
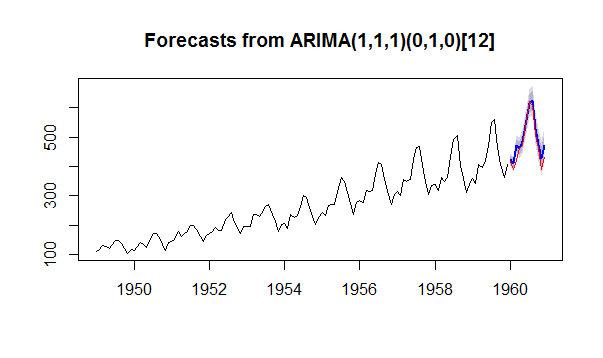
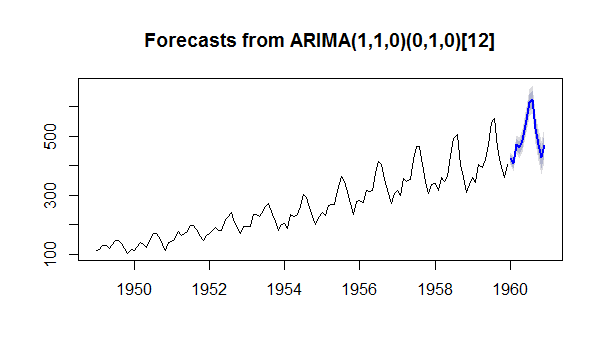
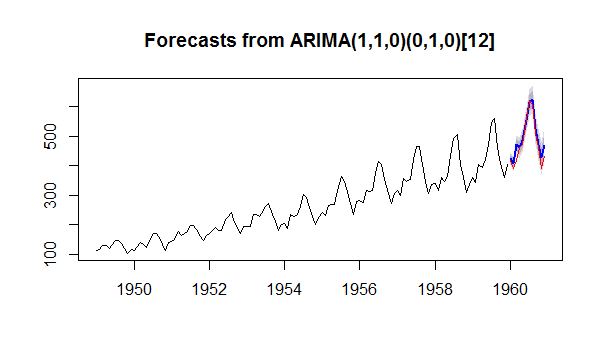
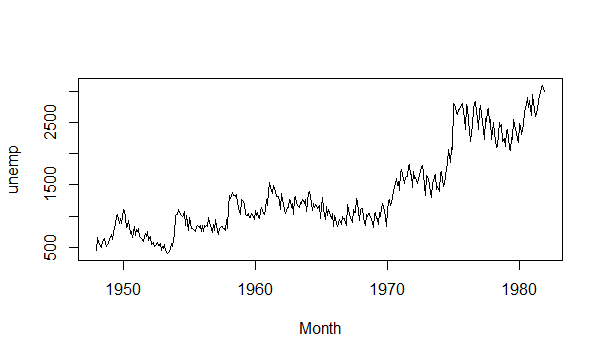
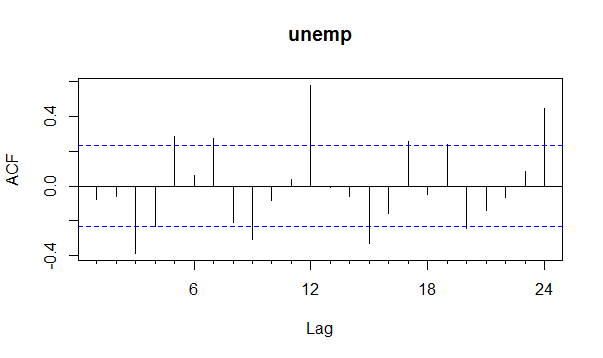
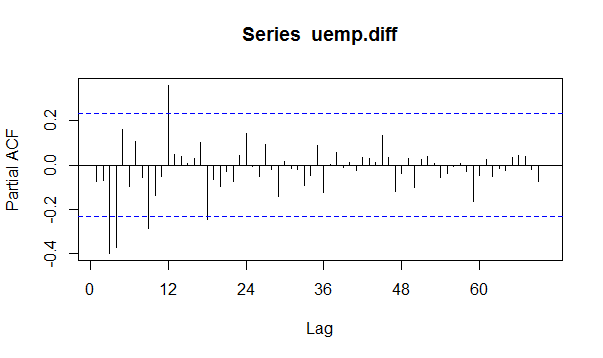
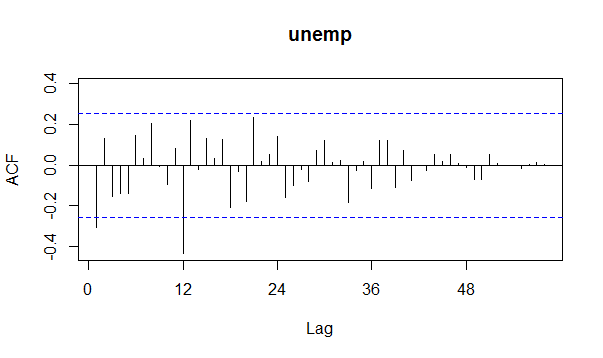
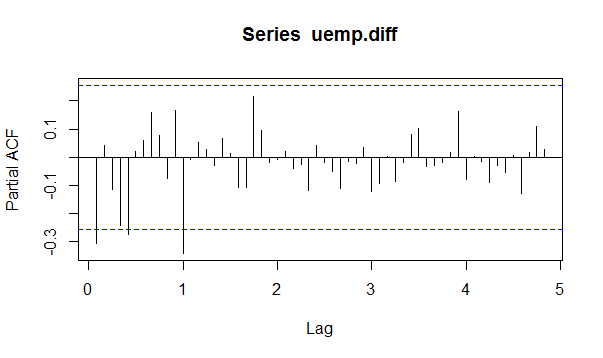
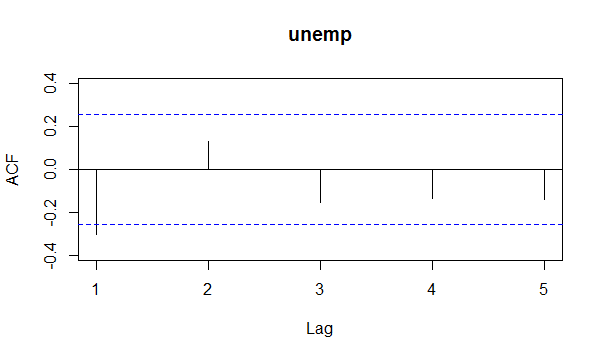
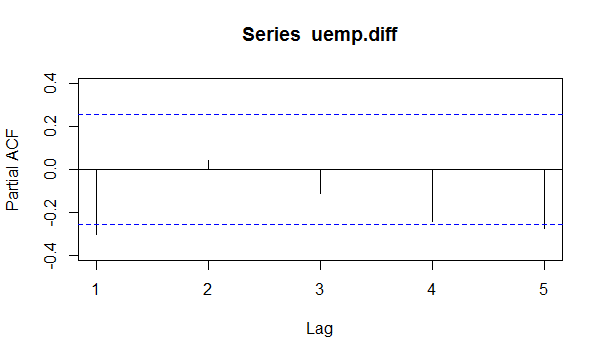
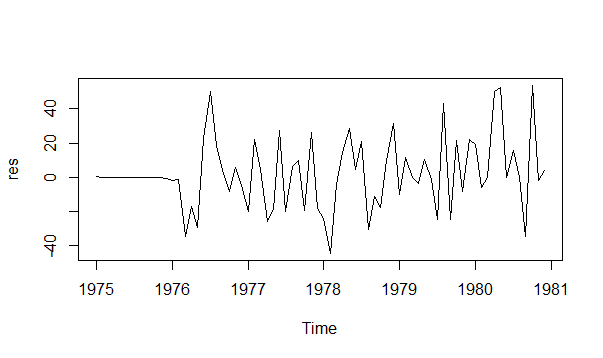
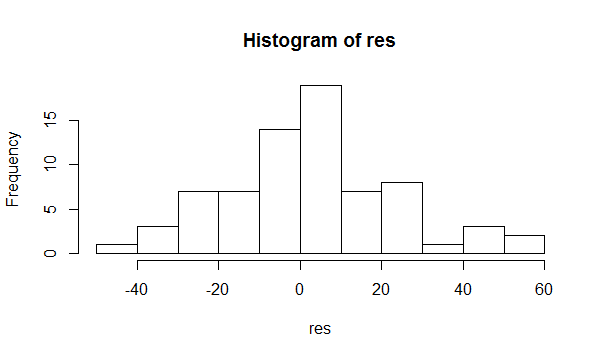
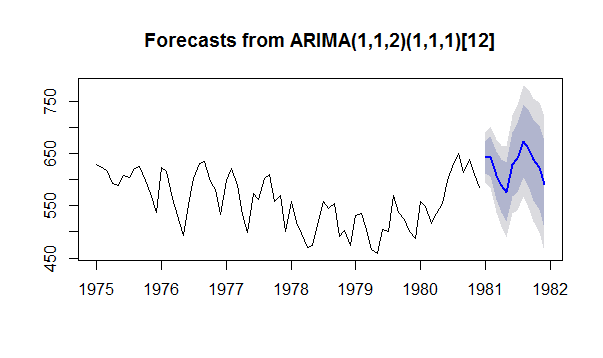
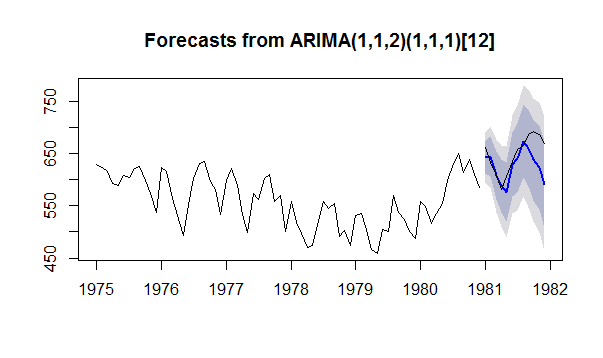
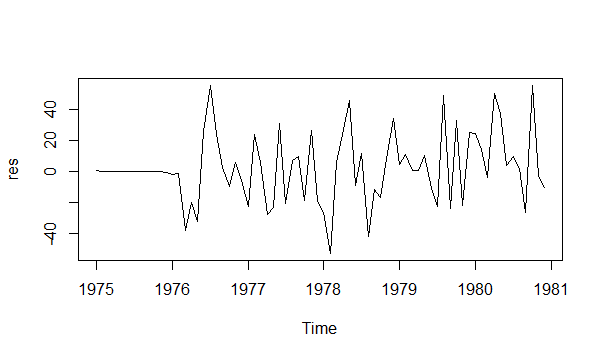
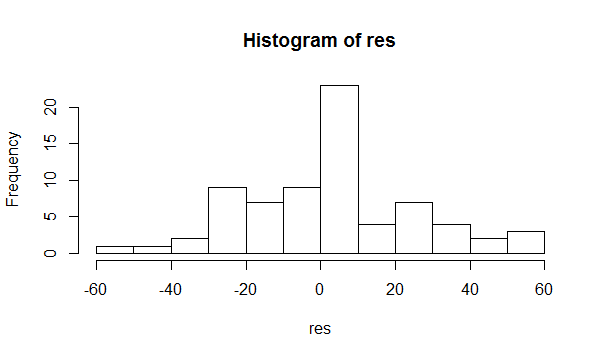
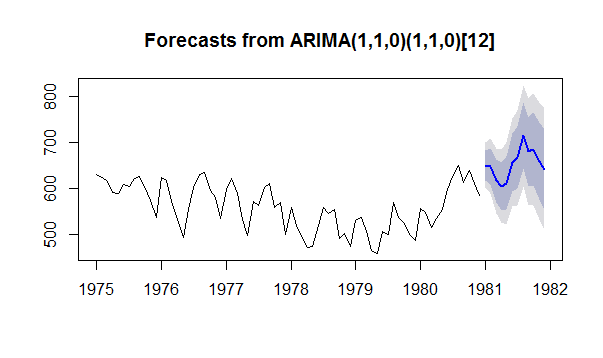
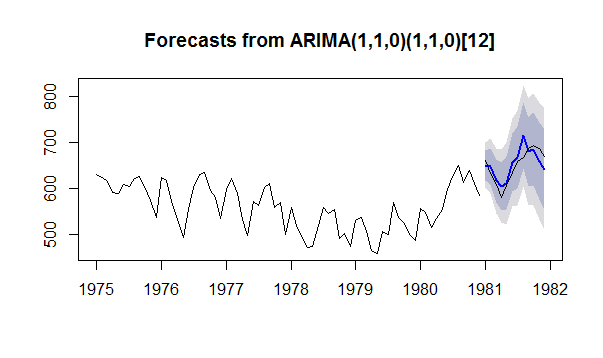
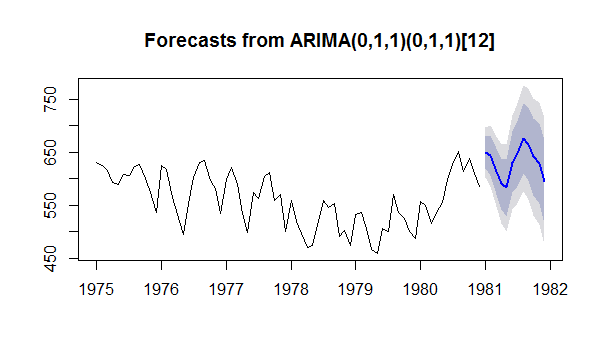
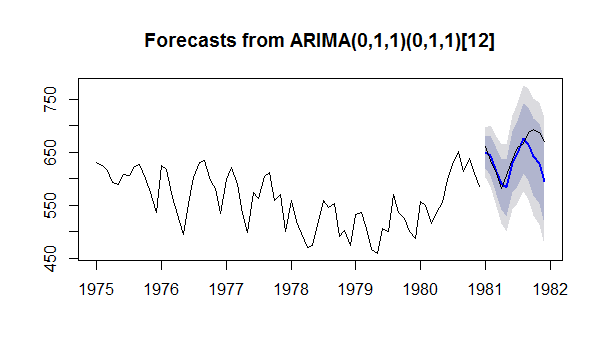
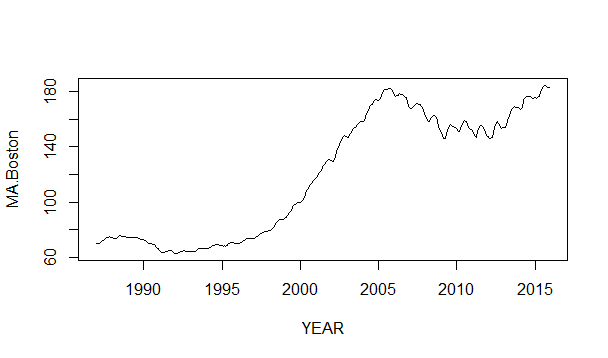
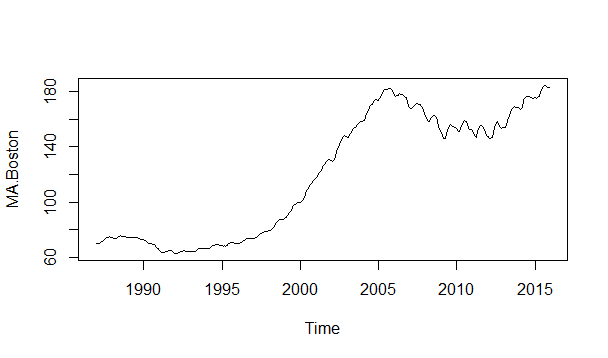
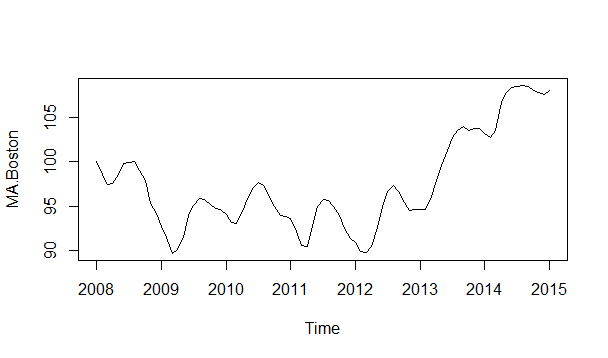
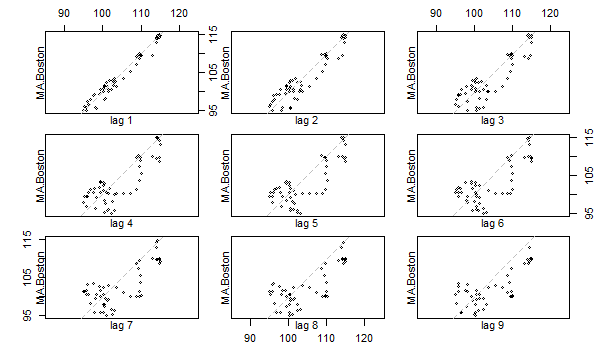
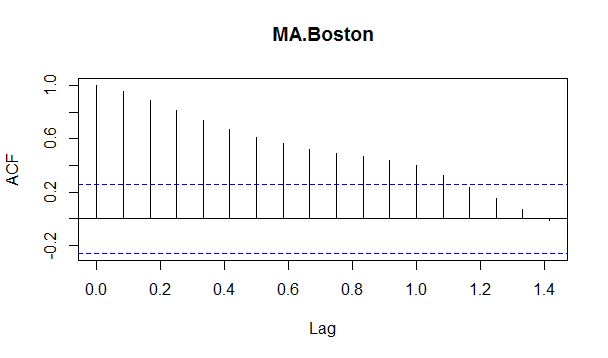
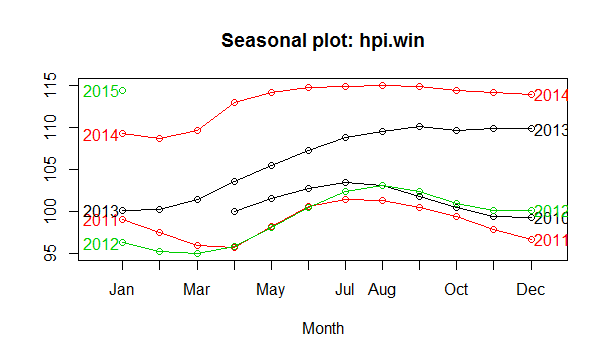
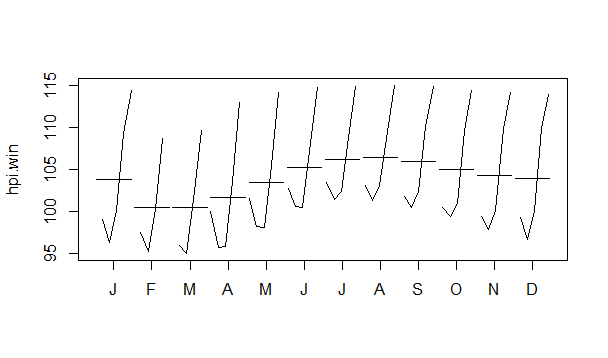
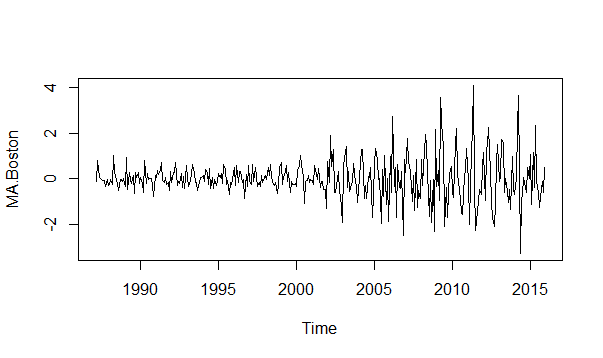
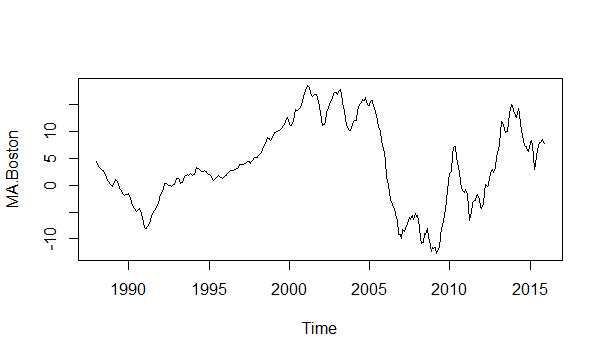
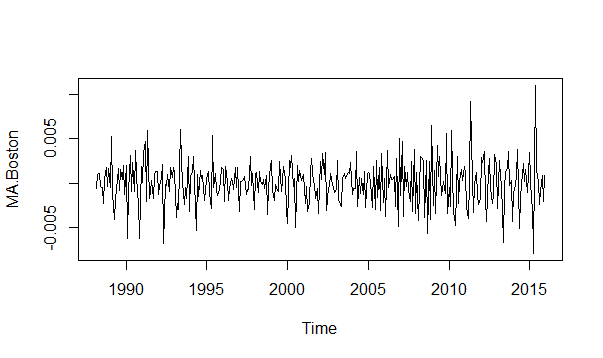
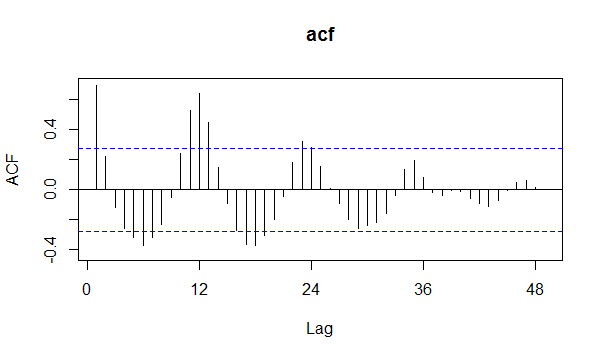
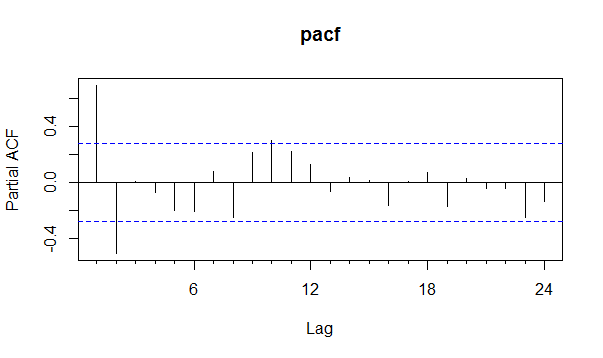
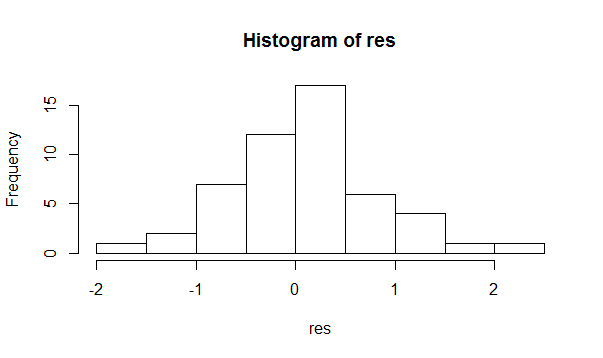
fc=forecast(fit,h=12)

plot(fc)

lines(uemp.test)

accuracy(fc,uemp.test)

OUTPUT:



Augmented Dickey-Fuller Test

data: hpi.diff

Dickey-Fuller = -11.039, Lag order = 6, p-value = 0.01

alternative hypothesis: stationary

Warning message:

In adf.test(hpi.diff, alternative = "stationary") :

p-value smaller than printed p-value

KPSS Test for Level Stationarity

data: hpi.diff

KPSS Level = 0.010357, Truncation lag parameter = 4, p-value = 0.1

Warning message:

In kpss.test(hpi.diff) : p-value greater than printed p-value

KPSS Test for Level Stationarity

data: hpi.diff

KPSS Level = 0.28147, Truncation lag parameter = 1, p-value = 0.1

Warning message:

In kpss.test(hpi.diff) : p-value greater than printed p-value

> adf.test(hpi.diff)

Augmented Dickey-Fuller Test

data: hpi.diff

Dickey-Fuller = -3.9789, Lag order = 3, p-value = 0.01766

alternative hypothesis: stationary

> fit

Series: hpi.train

ARIMA(2,1,1)

Coefficients:

ar1 ar2 ma1

1.0462 -0.4832 0.0386

s.e. 0.2994 0.2386 0.3632

sigma^2 estimated as 0.6289: log likelihood=-58.45

AIC=124.9 AICc=125.79 BIC=132.55

> Box.test(res,lag=10,fitdf=3,type="L")

Box-Ljung test

data: res

X-squared = 8.0697, df = 7, p-value = 0.3265

> accuracy(fc,hpi.test)

ME RMSE MAE MPE MAPE MASE ACF1

Training set 0.1182871 0.7613086 0.5733328 0.1116476 0.5552833 0.12748667 -0.01621305

Test set 0.1374146 0.2720267 0.2180882 0.1194609 0.1902554 0.04849423 0.27221634

Theil's U

Training set NA

Test set 0.893441

> fit\_auto

Series: hpi.train

ARIMA(0,1,1)(0,1,0)[12]

Coefficients:

ma1

0.7251

s.e. 0.1552

sigma^2 estimated as 0.4222: log likelihood=-37.23

AIC=78.46 AICc=78.8 BIC=81.74

> accuracy(fc\_auto,hpi.test)

ME RMSE MAE MPE MAPE MASE

Training set 0.1243748 0.5534702 0.4074154 0.1247289 0.3931342 0.09059317

Test set -2.3373127 2.4243299 2.3373127 -2.0425849 2.0425849 0.51972643

ACF1 Theil's U

Training set -0.04020716 NA

Test set 0.38518518 7.856406

|  |
| --- |
| > air.train  Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec  1949 112 118 132 129 121 135 148 148 136 119 104 118  1950 115 126 141 135 125 149 170 170 158 133 114 140  1951 145 150 178 163 172 178 199 199 184 162 146 166  1952 171 180 193 181 183 218 230 242 209 191 172 194  1953 196 196 236 235 229 243 264 272 237 211 180 201  1954 204 188 235 227 234 264 302 293 259 229 203 229  1955 242 233 267 269 270 315 364 347 312 274 237 278  1956 284 277 317 313 318 374 413 405 355 306 271 306  1957 315 301 356 348 355 422 465 467 404 347 305 336  1958 340 318 362 348 363 435 491 505 404 359 310 337  1959 360 342 406 396 420 472 548 559 463 407 362 405 |
|  |
| |  | | --- | | > | |

> air.test

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

1960 417 391 419 461 472 535 622 606 508 461 390 432

> kpss.test(air.diff)

KPSS Test for Level Stationarity

data: air.diff

KPSS Level = 0.043228, Truncation lag parameter = 2, p-value = 0.1

> adf.test(air.diff)

Augmented Dickey-Fuller Test

data: air.diff

Dickey-Fuller = -5.4971, Lag order = 4, p-value = 0.01

alternative hypothesis: stationary

Warning message:

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

fit

Series: air.train

ARIMA(1,1,1)(0,1,0)[12]

Coefficients:

ar1 ma1

-0.5528 0.3317

s.e. 0.2402 0.2693

sigma^2 estimated as 109.9: log likelihood=-447.49

AIC=900.97 AICc=901.18 BIC=909.31

> Box.test(res, lag=24, fitdf=2,type="L")

Box-Ljung test

data: res

X-squared = 27.945, df = 22, p-value = 0.1775

.

> accuracy(fc,air.test)

ME RMSE MAE MPE MAPE MASE

Training set 0.5549277 9.867979 7.417113 0.1165406 2.860269 0.2435834

Test set -18.2416015 24.806679 19.558125 -4.2033004 4.414960 0.6423029

ACF1 Theil's U

Training set -0.001923895 NA

Test set 0.054511577 0.5511927

> fit\_auto

Series: air.train

ARIMA(1,1,0)(0,1,0)[12]

Coefficients:

ar1

-0.2431

s.e. 0.0894

sigma^2 estimated as 109.8: log likelihood=-447.95

AIC=899.9 AICc=900.01 BIC=905.46

> accuracy(fc\_auto,air.test)

ME RMSE MAE MPE MAPE MASE ACF1

Training set 0.579486 9.907267 7.483159 0.1187348 2.880429 0.2457523 0.01227544

Test set -16.986385 23.931703 18.527682 -3.9334909 4.182395 0.6084625 0.04802038

Theil's U

Training set NA

Test set 0.5336134

> uemp.train

Jan Feb Mar Apr May Jun Jul Aug Sep

1975 629.8206 623.9910 616.1435 592.3767 589.2377 608.5202 605.1570 622.4215 626.6816

1976 624.2152 616.8161 565.2466 531.8386 493.7220 553.5874 602.2422 631.1659 634.7534

1977 598.6547 621.3004 589.2377 533.8565 498.6547 573.0942 563.2287 603.1390 611.2108

1978 559.1928 519.0583 497.0852 470.8520 474.4395 519.9552 558.5202 545.2915 553.8117

1979 531.3901 536.3229 505.6054 465.6951 458.9686 505.6054 500.6726 569.2825 536.7713

1980 557.6233 549.1031 515.6951 535.2018 554.7085 597.9821 625.7848 651.1211 613.6771

Oct Nov Dec

1975 599.3274 573.5426 536.7713

1976 596.6368 580.7175 534.3049

1977 558.9686 570.4036 500.4484

1978 491.2556 502.4664 474.6637

1979 524.8879 500.2242 486.7713

1980 638.7892 610.5381 585.8744

> uemp.test

Jan Feb Mar Apr May Jun Jul Aug Sep

1981 661.4350 633.4081 609.1928 581.3901 606.0538 635.8744 658.7444 667.0404 686.9955

Oct Nov Dec

1981 693.2735 686.7713 670.6278

> adf.test(uemp.train)

Augmented Dickey-Fuller Test

data: uemp.train

Dickey-Fuller = -0.64416, Lag order = 4, p-value = 0.9709

alternative hypothesis: stationary

> adf.test(uemp.diff)

Augmented Dickey-Fuller Test

data: uemp.diff

Dickey-Fuller = -4.8856, Lag order = 4, p-value = 0.01

alternative hypothesis: stationary

> kpss.test(uemp.diff)

KPSS Test for Level Stationarity

data: uemp.diff

KPSS Level = 0.05199, Truncation lag parameter = 1, p-value = 0.1

> Box.test(res, lag=24, fitdf=6,type="L")

Box-Ljung test

data: res

X-squared = 16.278, df = 18, p-value = 0.5732

> accuracy(fc,uemp.test)

ME RMSE MAE MPE MAPE MASE ACF1

Training set 2.017079 20.84085 14.94247 0.3265388 2.729260 0.3791585 -0.006002043

Test set 22.429710 36.10915 26.78700 3.3249928 4.020289 0.6797080 0.523971426

Theil's U

Training set NA

Test set 1.62452

|  |
| --- |
| > Box.test(res, lag=12, fitdf=6,type="L")  Box-Ljung test  data: res  X-squared = 8.6215, df = 6, p-value = 0.196 |
|  |
| |  | | --- | | > | |

> lines(uemp.test)

> accuracy(fc,uemp.test)

ME RMSE MAE MPE MAPE MASE ACF1

Training set 2.124234 22.69808 16.62803 0.3518397 3.054613 0.4219287 0.01043188

Test set -4.034066 21.47913 17.59611 -0.7055019 2.706010 0.4464935 0.24530696

Theil's U

Training set NA

Test set 0.998522

|  |
| --- |
| > accuracy(fc,uemp.test)  ME RMSE MAE MPE MAPE MASE ACF1  Training set 2.203459 20.75460 15.00709 0.3520972 2.742570 0.3807982 -0.02292445  Test set 19.194422 33.79398 24.49976 2.8269864 3.677646 0.6216703 0.57030347  Theil's U  Training set NA  Test set 1.522734 |
|  |
| |  | | --- | | > | |