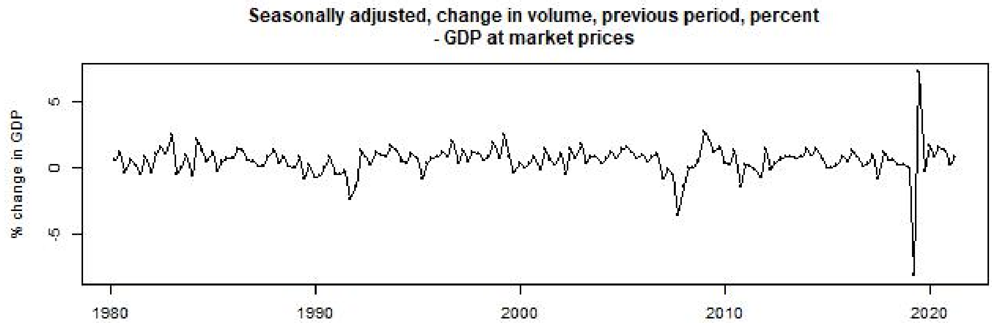
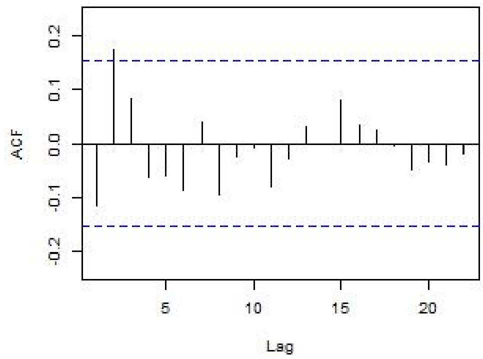
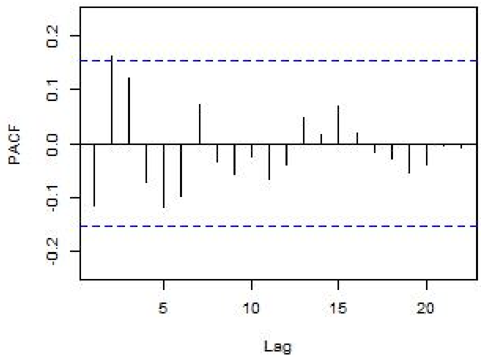
Assignment 2 | Jack (Quan Cheng) Xie
Professor Yukai Yang | 20 September 2022
Time Series Econometrics

Problem 1.

The time series, auto correlation function, and partial auto correlation function plots are below.







Based on the dampened oscillation in both ACF and PACF and the spikes above the significance thresholds at lag 2, I choose ARMA(2, 2).

Problem 2.

Results of the table are as below.

Model	AIC	h	RMSE(h)	MAE(h)	BIAS(h)
ARMA(p=1, q=1)	26.546	1	0.547	0.419	-0.118
		2	0.463	0.392	-0.058
ARMA(p=2, q=1)	24.387	1	0.54	0.421	-0.112
		2	0.457	0.383	-0.043
ARMA(p=1, q=2)	26.356	1	0.52	0.414	-0.094
		2	0.451	0.364	-0.015
ARMA(p=2, q=2)	24.057	1	0.497	0.405	-0.091
		2	0.446	0.362	-0.046
choice.ARMA(p=2, q=2)	24.057	1	0.497	0.405	-0.091
		2	0.446	0.362	-0.046
auto.ARMA(p=0, q=3)	25.86	1	0.536	0.415	-0.085
		2	0.422	0.368	-0.035

For the $h = 1$ step forecast, the ARMA(2, 2) model has lower errors than any other model, and has less bias except for ARMA(0, 3) = MA(3).

For the $h = 2$ step forecast, the MA(3) has less RMSE and bias, but ARMA(2, 2) has less MAE.

Appendix. Code also available in `.R` attachment.

```
library(rio)
library(dplyr)
library(magrittr)
library(forecast)
library(xtable)

### Load and clean data

# https://scb.se/en/finding-statistics/statistics-by-subject-area/national-accounts/national-account
# 11. GDP: expenditure approach (ENS 2010), seasonally adjusted (xlsx)
url <- "https://www.statistikdatabasen.scb.se/sq/17384"

raw <- rio::import(url)

# keep only row row number
n = 33
title = paste(raw[n, 1], raw[n, 2], sep='\n')
gdp <- raw[n, ] %>% t()

# set column name
colnames(gdp) <- title
rownames(gdp) <- raw[2, ]

# change in GDP volume
gdp <- gdp[-c(1:3)] %>%
  replace('..', 'na') %>%
  as.numeric() %>%
  na.omit() %>%
  ts(start=c(1980, 2), frequency = 4)


### Questions

### Question 1.1

# plot trend data, ACF and PACF
jpeg("HW2-gdp_tsdisplay.jpg", width = 650, height = 500)
tsdisplay(gdp, main=title, ylab='% change in GDP')
dev.off()

### Question 1.2

## generate prediction

model_predict <- function(model, sample, h=1, frequency=4){

  modelorder = arimaorder(model)

  # start end time of sample
  test_start <- time(sample)[1]
  test_end <- time(sample)[length(sample)]

  preds <- ts(NA, start=test_start+(h-1)/frequency,
              end=test_end,
              frequency=frequency)
```

```
  k <- length(preds)

  for (t in 1:k) {
    loopmodel <- Arima(window(gdp,end=test_start + (h+t-3)/frequency),
                       order=modelorder)

    preds[t] <- forecast(loopmodel, h=h)$mean[h]
  }

  return(preds)
}


## evaluation functions

RMSE <- function(model, sample, h=1){

  pred = model_predict(model, sample, h=h)

  k = length(pred)
  res <- 0
  for (t in 1:k) {
    res = res + (pred[t] - sample[t+h-1])^2
  }
  res = (res/(k-h+1))^(1/2)
  return(res)
}

MAE <- function(model, sample, h=1){
  pred = model_predict(model, sample, h=h)
  k = length(pred)
  res <- 0
  for (t in 1:k) {
    res = res + abs(pred[t] - sample[t+h-1])
  }
  res = (res/(k-h+1))
  return(res)
}

bias <- function(model, sample, h=1){
  pred = model_predict(model, sample, h=h)
  k = length(pred)
  res <- 0
  for (t in 1:k) {
    res = res + pred[t] - sample[t+h-1]
  }
  res = (res/(k-h+1))
  return(res)
}


# evaluation create models

orders <- matrix(NA, nrow=6, ncol=3)
ordnames <- matrix(NA, nrow=nrow(orders))
i <- 0
for (q in 1:2) {
  for (p in 1:2) {
```

```
    i = i + 1
    ordnames[i] = 'ARMA'
    orders[i, ] = c(p, 0, q)
  }
}

# self chosen model
ordnames[5] = 'choice.ARMA'
orders[5,] = c(2, 0, 2)

# auto model
model <- auto.arima(gdp, seasonal=FALSE, max.p=4, max.q=4)
modelorder = arimaorder(model)

ordnames[6] = 'auto.ARMA'
orders[6,] = modelorder

ordnames = paste(ordnames, '(p=', orders[,1], ', q=', orders[,3],')', sep='')
rownames(orders) = ordnames
orders

## evaluate the models on sample

sample = window(gdp, start=c(2012, 1), end = c(2015, 2))
res = matrix('', nrow=2*nrow(orders), ncol=6)

n <- 1

for (i in 1:nrow(orders)){

  cat('\n--------------------------------\n')

  name <- ordnames[i]
  cat(name, '\n')

  model <- arima(sample, order = orders[i, ])
  # plot(model)

  aic_val = AIC(model)
  cat('AIC:', aic_val, '\n\n')

  res[n, 1] = name
  res[n, 2] = round(aic_val, 3)

  for (h in 1:2) {
    rmse_val = RMSE(model, sample, h=h)
    mae_val = MAE(model, sample, h=h)
    bias_val = bias(model, sample, h=h)

    cat(gsub('%h', h, 'RMSE(h = %h) ='), rmse_val, ' ')
    cat(gsub('%h', h, 'MAE(h = %h) ='), mae_val, ' ')
    cat(gsub('%h', h, 'bias(h = %h) ='), bias_val, '\n')

    res[n, 3] = paste(h)
    res[n, 4] = round(rmse_val, 3)
    res[n, 5] = round(mae_val, 3)
    res[n, 6] = round(bias_val, 3)
```

```
    n = n + 1

  }
}

colnames(res) <- c('Model', 'AIC', 'h',
                   'RMSE(h)', 'MAE(h)', 'BIAS(h)'
                   )

tab = xtable(res, type = "latex")

align(tab) <- "c|l|c||c|ccc|"

hlines <- append(c(-1, -1), 2 * c(0:6))
hlines <- append(hlines, 12)

print(tab, file = "HW2-res_table.tex",
      hline.after = hlines,
      include.rownames = FALSE
      )
```

Time Series Econometrics: Home work assignment 2

Yukai Yang
Department of Statistics
Uppsala University

Abstract

Please write your report in LaTeX. The report should be clearly written such that it is easy to understand what is done and why. Please attach any computer code in an appendix.

1 Forecast Evaluation

Go to `http://scb.se/en/`. Download data for the change in volume of seasonally adjusted GDP.

1. Plot the time series with proper labeling of axes and dates. Also plot the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the series. Based on these plots, make a guess of an ARMA(p, q) model that you believe could be appropriate.

2. Perform a forecast evaluation of one- and two-step predictions, where the models to be evaluated are:

   (a) ARMA$(1, 1)$, ARMA$(2, 1)$, ARMA$(1, 2)$ and ARMA$(2, 2)$

   (b) Your choice of ARMA(p, q) in the previous part

   (c) The "best" ARMA according to `auto.arima()` in the `forecast` package[1], with `max.p` and `max.q` set to 4

[1]If you are not using R, you may instead compute the AIC for all ARMA(p, q) with $p, q \leq 4$ and choose the best model according to this.

To perform the evaluation, let 2012Q1-2015Q2 be the evaluation part of the sample. Estimate the models on data up to and including 2011Q4 and make predictions for 2012Q1 and 2012Q2. Add the data point 2012Q1, re-estimate the models and make predictions for 2012Q2 and 2012Q3. Continue until you reach the end of the sample.

For each model j, you will have two sequences of forecasts: $\{\hat{y}^{(j)}_{t+1|t}\}_{t=2011Q4}^{2015Q1}$ and $\{\hat{y}^{(j)}_{t+2|t}\}_{t=2011Q4}^{2014Q4}$. To compare the forecasts, use root mean squared error (RMSE), mean absolute deviation (MAD) and bias:

$$RMSE^{(j)}_h = \sqrt{\sum_{t=2011Q4}^{2015Q2-h} \frac{\left(y_{t+h} - \hat{y}^{(j)}_{t+h|t}\right)^2}{\#\text{ forecasts}}}, \quad MAE^{(j)}_h = \sum_{t=2011Q4}^{2015Q2-h} \frac{\left|y_{t+h} - \hat{y}^{(j)}_{t+h|t}\right|}{\#\text{ forecasts}}$$

$$Bias^{(j)}_h = \sum_{t=2011Q4}^{2015Q2-h} \frac{y_{t+h} - \hat{y}^{(j)}_{t+h|t}}{\#\text{ forecasts}}.$$

Compare the models and try to answer which model is the better forecaster. Is there a unanimous winner? Do the three measures agree on which is better? Is there a difference between considering one- and two-step forecasts?