Homework 7

Time Series Analysis (STAT 6391)

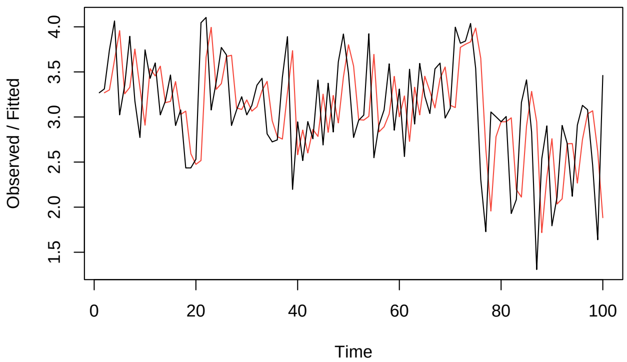
Willliam Ofosu Agyapong*

October 31, 2023

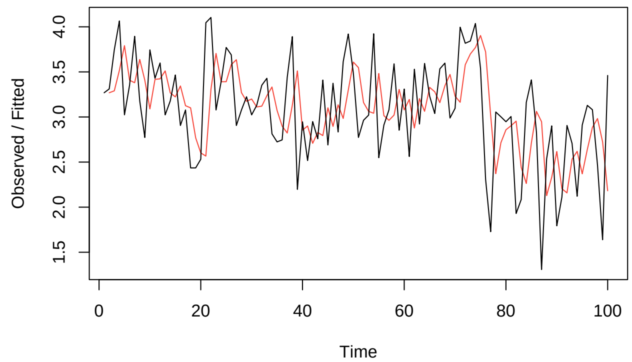
*woagyapong@miners.utep.edu, PhD Data Science, University of Texas at El Paso (UTEP).

Problem 5.1

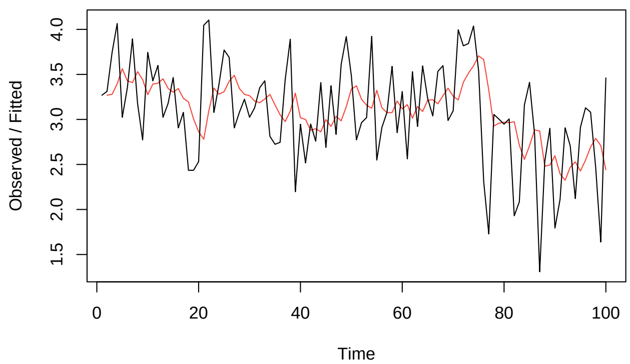
We calculate the EWMA, x_{n+1}^n, for the first 100 observations for the logarithm of the glacial varve data, say x_t with $\lambda = .25, .50$ and $.75$.



(a) $\lambda = .25$



(b) $\lambda = .50$



(c) $\lambda = .75$

Figure 1: Logarithm of the first 100 glacial varve data (in black) with EWMAs for $\lambda = .25, .50, .75$ superimposed (in red).

From the plots in Figure 1, we make the following observations:

- For $\lambda = .25$, the the logarithm of the data and the EWMA are indistinguishable (i.e., the forcasts by the EWMA is too wiggly).
- For $\lambda = .50$, the EWMA is less wiggly than that for $\lambda = .25$ but less smoother than that for $\lambda = .75$.
- For $\lambda = .75$, the EWMA is less wiggly and forms a smooth curve.

In general, we observe that, as the value of λ increases from .25 to .75, the EWMA becomes less wiggly and more smoother. Thus, confirming the general behavior of EWMAs that larger values of λ produce smoother forecasts (Shumway and Stoffer 2019).

Problem 5.2

Exploring the US GDP series

For this problem, we repeat the analysis done in Example 5.6 in (Shumway and Stoffer 2019) for the US GDP series. We start with a time series plot of the actual GDP series with a corresponding ACF plot as shown in 2. We observe a strong trend in the data with periodic dips. The ACF plot exhibits a slow decay, indicating that differencing may be needed, and hence affirming the appropriateness of analyzing the growth rate of GDP instead of the actual GDP.

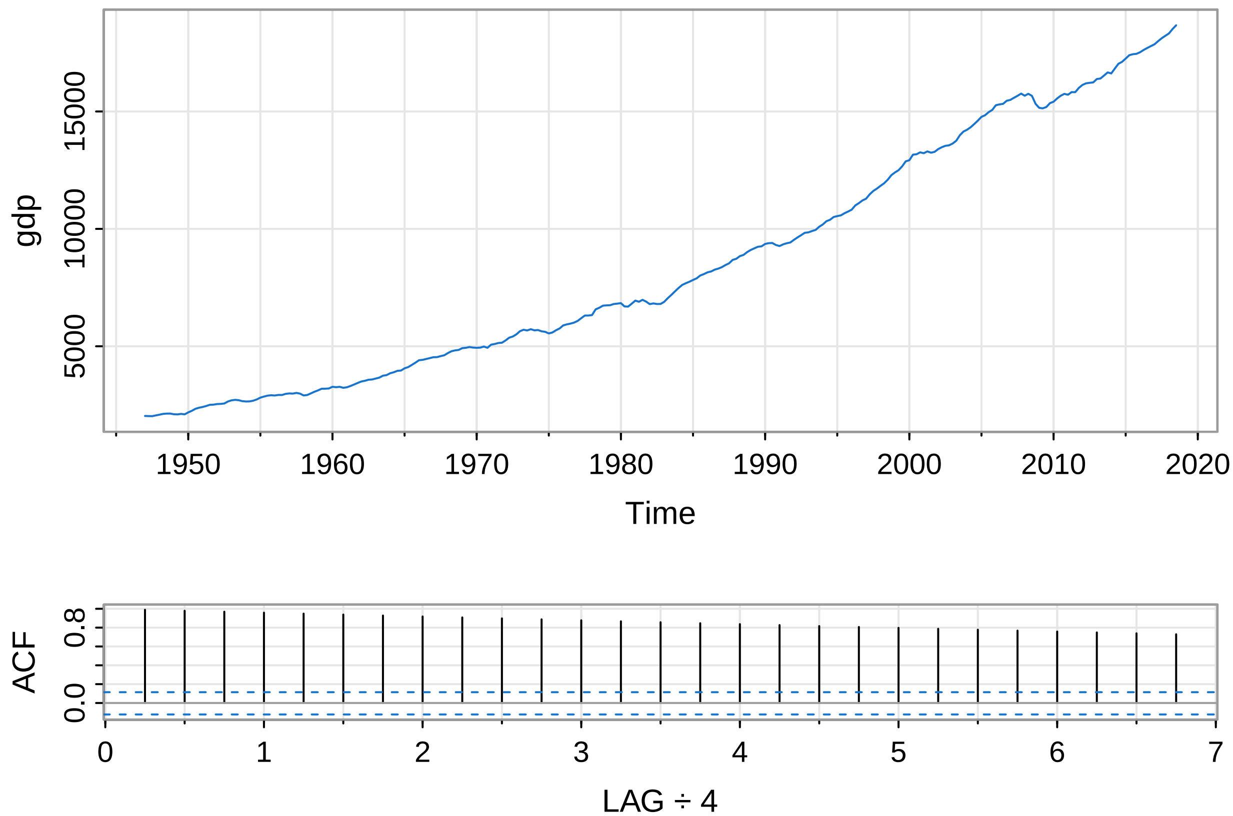
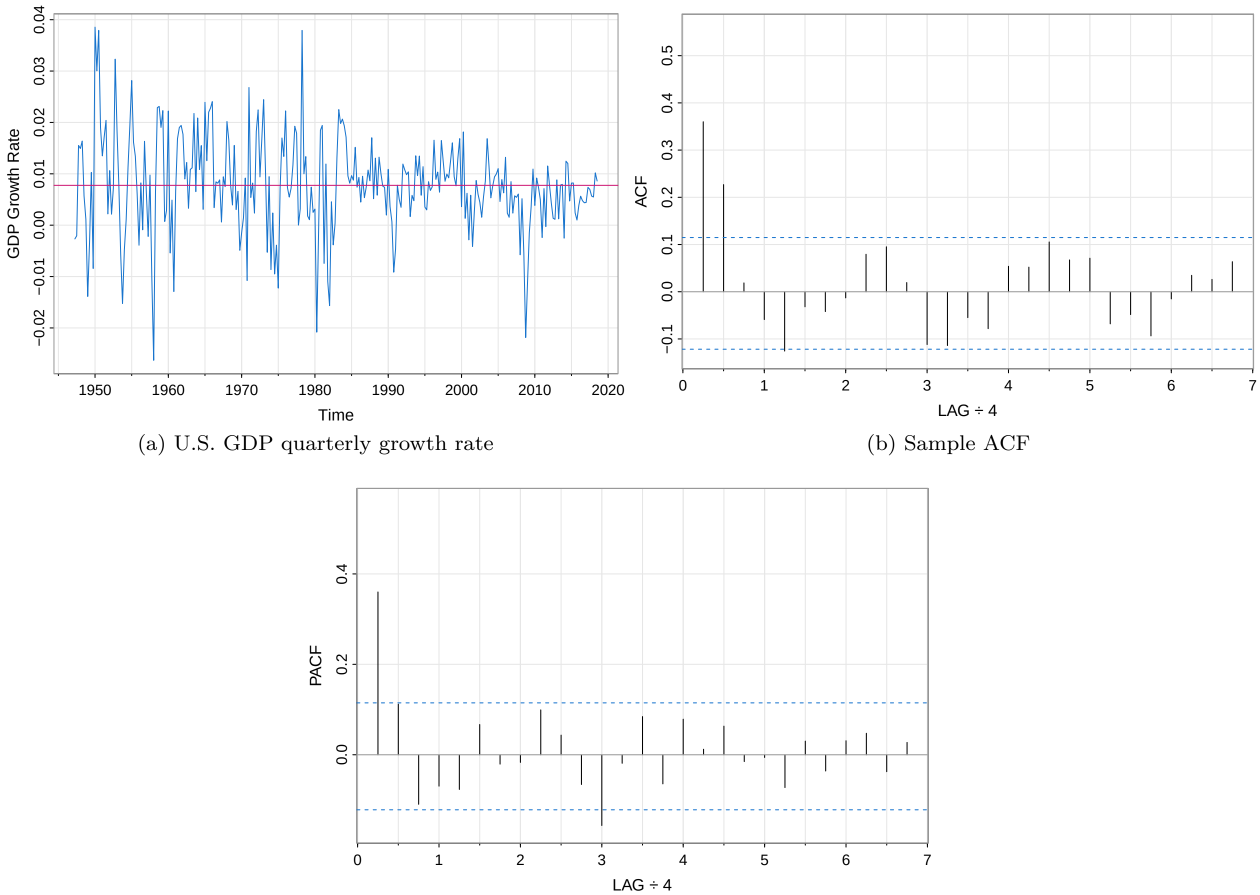


Figure 2: Top panel: Seasonally adjusted quarterly U.S. GDP from 1947(1) to 2018(3). Bottom panel: Sample ACF of the GDP data.

Following our previous observations, we present in Figure 3 a time series plot of the quarterly GDP growth rate along with the sample ACF and PACF of the growth rate. The quarterly GDP growth rate appears to be a stable process.



(a) U.S. GDP quarterly growth rate

(b) Sample ACF

(c) Sample PACF

Figure 3: U.S. GDP quarterly growth rate from 1947(1) to 2018(3) along with sample ACF and sample PACF. The horizontal red line displays the average growth of the process, which is approximately 1%. Lag is in years.

From Figure 3, we see that the sample ACF of GDP the growth rate appears to cut off at lag 2 and the PACF is somehow tailing off, suggesting that the GDP growth rate follows an MA(2) process, or log GDP follows an ARIMA(0,1,2) model. It also seems quite reasonable to also suggest that the ACF is tailing off while the PACF is cutting off at lag 1. This leads to two competing models, MA(2) and AR(1) for us to diagnose.

In the next section, we consider parameter estimates and diagnostics for the two models.

Parameter estimation for GDP Growth Rate

Parameter estimates for the suggested models for the GDP growth rate for MA(2) process and AR(1) process are presented in Tables 1 and 2, respectively.

All the regression coefficients are statistically significant, including the estimate for the constant term. We will defer comparison of the performance metrics to the last stage of the analysis where we choose the final model.

	Estimate	SE	t.value	p.value
Coefficients				
MA1	0.3070	0.0579	5.2988	0
MA2	0.2258	0.0547	4.1271	0
Constant	0.0077	0.0008	9.8631	0
Performance metrics				
Sigma^2	0.0001			
AIC	-6.6363			
AICc	-6.6360			
BIC	-6.5852			

Table 1: Parameter estimates and performance metrics for MA(2) model for the GDP growth rate.

	Estimate	SE	t.value	p.value
Coefficients				
AR1	0.3603	0.0551	6.5366	0
Constant	0.0077	0.0008	9.5915	0
Performance metrics				
Sigma^2	0.0001			
AIC	-6.6256			
AICc	-6.6255			
BIC	-6.5873			

Table 2: Parameter estimates and performance metrics for AR(1) model for the GDP growth rate.

Diagnostics for GDP Growth Rate

Our next step involves residual diagnostics. Relevant diagnostics plots are provided in Figures 4 and 5, for the MA(2) fit and AR(1) fit, respectively.

Inspection of the time plot of the standardized residuals in Figure 4 shows no obvious patterns. It is worthy to note that there may be outliers because a few standardized residuals exceed 3 standard deviations in magnitude. The impact of these potential outliers may not severe given that there are no values that are extremely large in magnitude.

For the normality assumption, the Q-Q plot of the residuals suggests no serious violation. Moreover, the ACF of the residuals exhibits no obvious departure from the model assumptions, suggesting that the residuals are white.

Next, we turn our attention to the Q-statistic. Some of the associated p-values for Ljung-Box statistic are above the .05 level line, whereas some fall on the line, so that the null hypothesis that the residuals are white may not be rejected.

```
initial  value -4.672758
iter   2 value -4.749239
iter   3 value -4.750696
iter   4 value -4.750723
iter   5 value -4.750724
iter   6 value -4.750725
iter   7 value -4.750725
iter   7 value -4.750725
iter   7 value -4.750725
final  value -4.750725
```

```
converged
initial  value -4.751078
iter   2 value -4.751080
iter   3 value -4.751080
iter   4 value -4.751081
iter   5 value -4.751081
iter   5 value -4.751081
iter   5 value -4.751081
final  value -4.751081
converged
```

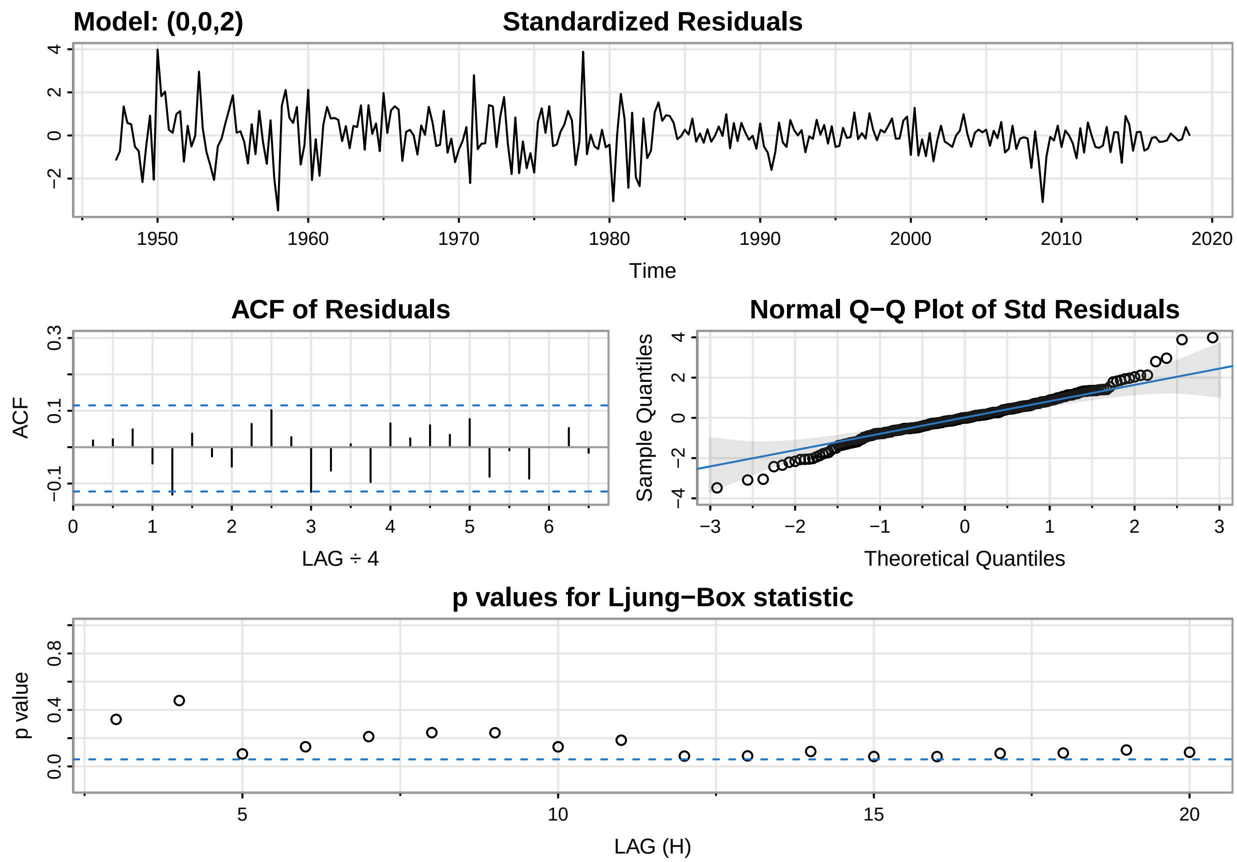


Figure 4: Diagnostics of the residuals from MA(2) fit on GDP growth rate.

Based on 5, analysis of the AR(1) residuals reveal similar outcomes, except that the corresponding p-values for Ljung-Box statistic provides less evidence of white noise of the residuals (because all p-values fall on the .05 level reference line) compared to the one from the MA(2) model.

```
initial  value -4.673186
iter   2 value -4.742918
iter   3 value -4.742921
iter   4 value -4.742923
iter   5 value -4.742925
iter   6 value -4.742925
iter   6 value -4.742925
final  value -4.742925
converged
```

```
initial  value -4.742229
iter   2 value -4.742234
iter   3 value -4.742245
iter   3 value -4.742245
iter   3 value -4.742245
final  value -4.742245
converged
```

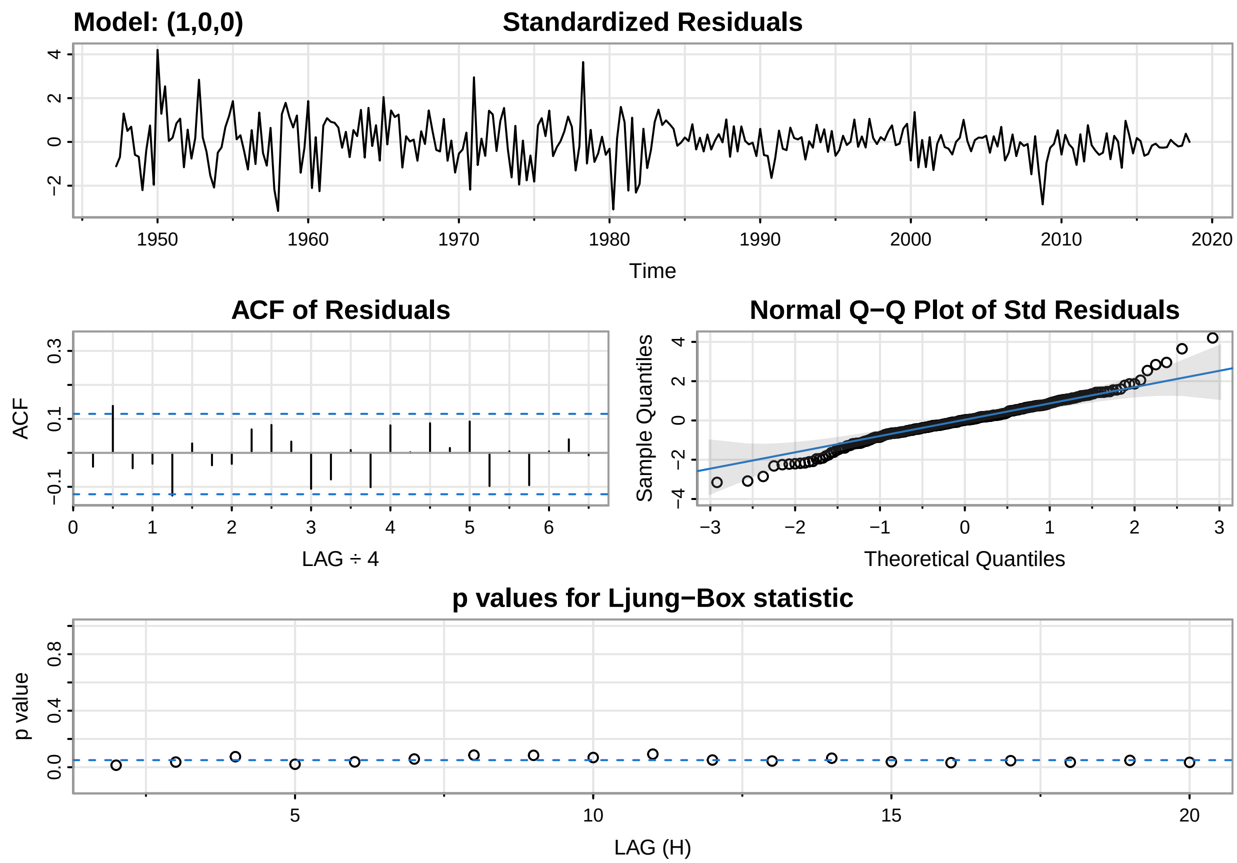


Figure 5: Diagnostics of the residuals from AR(1) fit on GDP growth rate.

After the diagnostics presented above, we attempted to overfit the models by trying other models including MA(3), MA(4), AR(2), and AR(3), to see if the results change significantly. Results of some of these complex models are presented in the Appendix, from which we concluded that the extra parameters do not significantly change the results.

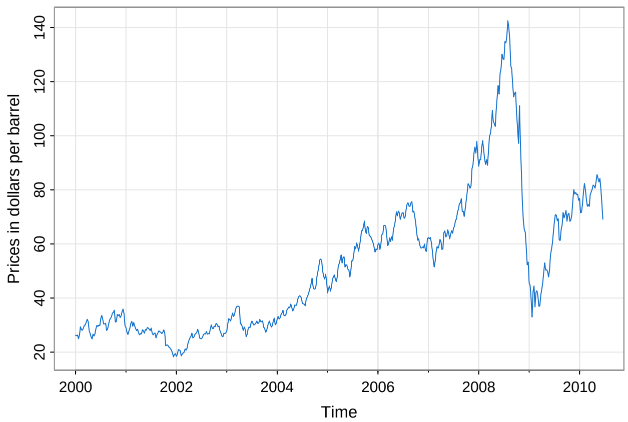
Final Model Choice for the U.S. GDP Series

Given the similarity in two models, we

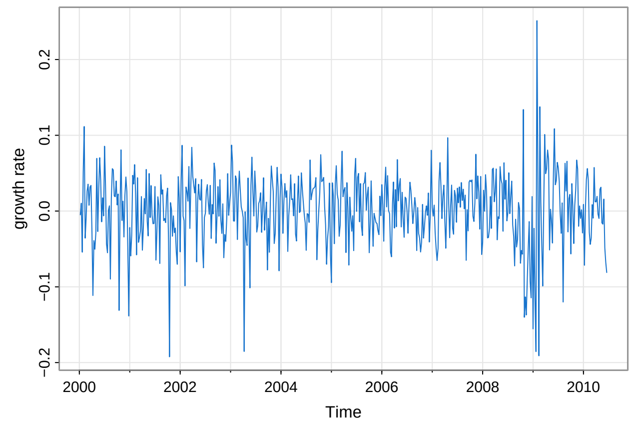
In this final stage, we refer back to performance metrics (AIC, AICs, and BIC) in Tables 1 and 2. It is clear from the results that the MA(2) performs better than the AR(1) model when comparing the AIC and AICs, while the BIC prefers the simpler AR(1) model. There is not a clear winner, however, for parsimony and the fact that AR models are easy to work with, we will conclude with AR(1) model for the GDP growth rate.

Problem 5.3

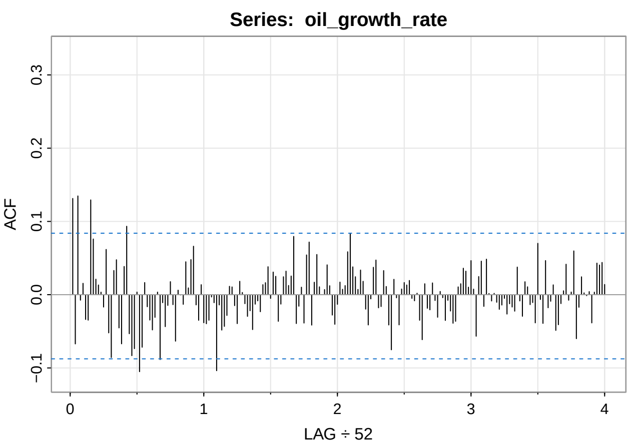
We begin by performing some exploratory analysis to provide a fair idea of the data to help us identify preliminary values of the AR order, p, and the MA order, q.



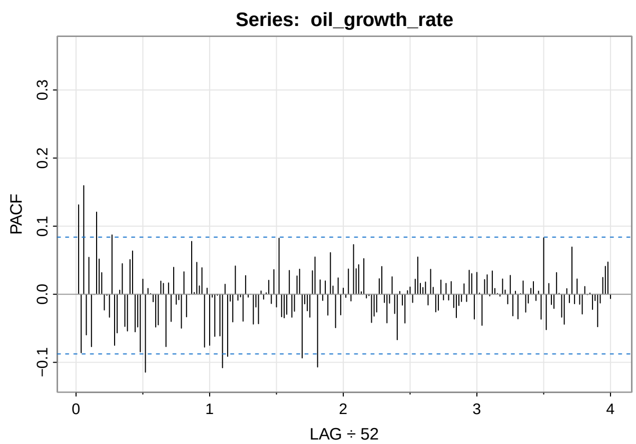
(a) Crude oil preices in dollars per barrel



(b) Growth rate of crude oil preices in dollars per barrel



(c) Sample ACF of the growth rate



(d) Sample PACF of the growth rate

Figure 6: Exploring the Crude oil preices in dollars per barrel series.

Time series plot of the actual crude oil prices per barrel shows strong trend in the data. Although the growth rate of oil prices per barrel appears stabilized, there is still some appearance of seasonal correlations.

Both the ACF and PACF are tailing off, which suggests that an ARIMA model may be appropriate. For starters, we fitted ARIMA(1,0,1) or ARMA(1,1) to the growth rate and later explored other forms of the ARIMA model. Most of the model choices with respect to d=0 and varying values of p and q produced somewhat identical models to ARIMA(1,0,1). However, ARIMA(0,0,3) appeared better among them all, so for brevity we presented diagnostics for only ARIMA(1,0,1) fit along with ARIMA(0,0,3) fit to the growth rate.

```
initial  value -3.057594
iter   2 value -3.061420
iter   3 value -3.067360
iter   4 value -3.067479
iter   5 value -3.071834
iter   6 value -3.074359
```

```
iter   7 value -3.074843
iter   8 value -3.076656
iter   9 value -3.080467
iter  10 value -3.081546
iter  11 value -3.081603
iter  12 value -3.081615
iter  13 value -3.081642
iter  14 value -3.081643
iter  14 value -3.081643
iter  14 value -3.081643
final  value -3.081643
converged
initial  value -3.082345
iter   2 value -3.082345
iter   3 value -3.082346
iter   4 value -3.082346
iter   5 value -3.082346
iter   5 value -3.082346
iter   5 value -3.082346
final  value -3.082346
converged
```

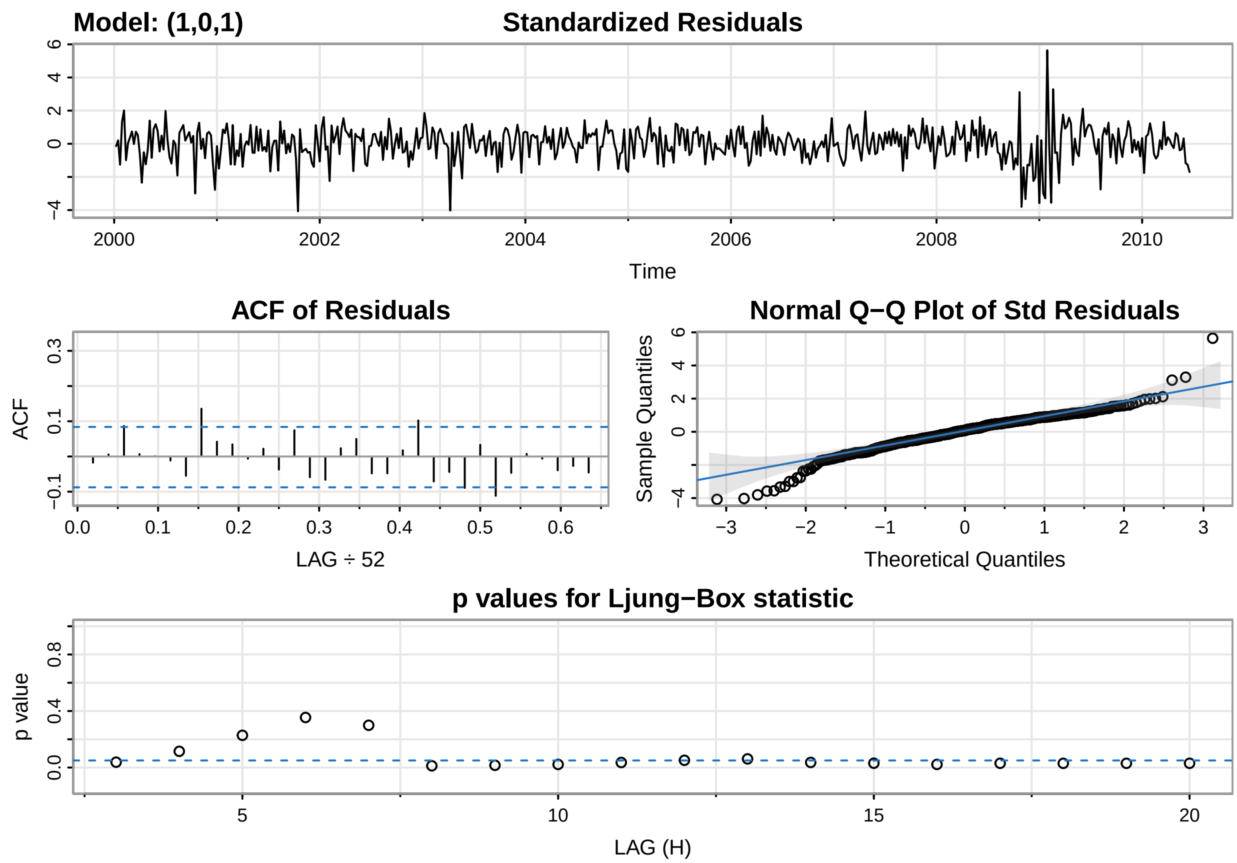


Figure 7: Diagnostics of the residuals from ARIMA(1,0, 1) fit on crude oil prices per barrel growth rate.

Inspection of the time plot of the standardized residuals in Figure 7 reveals no obvious patterns. However, some relatively large outliers are present. The ACF of the residuals exhibits mild departure (not too concerning though) from the model assumptions because some of the values extend slightly beyond the confidence bands. The Q-Q plot of the residuals suggests that the distribution of the residuals is reasonably normal.

Most of the p-values fall on the .05 horizontal line while 4 of them exceed .05, with no values falling far from the referenced line. This suggest that we have sufficient evidence (not too strong) not to reject the null hypothesis that the residuals are white.

Next, we present diagnostics of the residuals from ARIMA(0,0,3) fit to the growth rate of crude oil prices per barrel in Figure 8. The results look very identical to the results obtained in Figure 7 for the ARIMA(1,1) fit. The only obvious difference lies between the p-values for the Ljung-Box statistic plots for the two models, where the p-values for the ARIMA(0,0,3) fit provide stronger evidence of the residuals being white than the evidence provided by the ARIMA(1,0,1) fit.

```
initial  value -3.058495
iter   2 value -3.086110
iter   3 value -3.086980
iter   4 value -3.087501
iter   5 value -3.087521
iter   6 value -3.087521
iter   7 value -3.087522
iter   8 value -3.087522
iter   9 value -3.087522
iter   9 value -3.087522
iter   9 value -3.087522
final  value -3.087522
converged
initial  value -3.087448
iter   2 value -3.087448
iter   3 value -3.087449
iter   3 value -3.087449
iter   3 value -3.087449
final  value -3.087449
converged
```

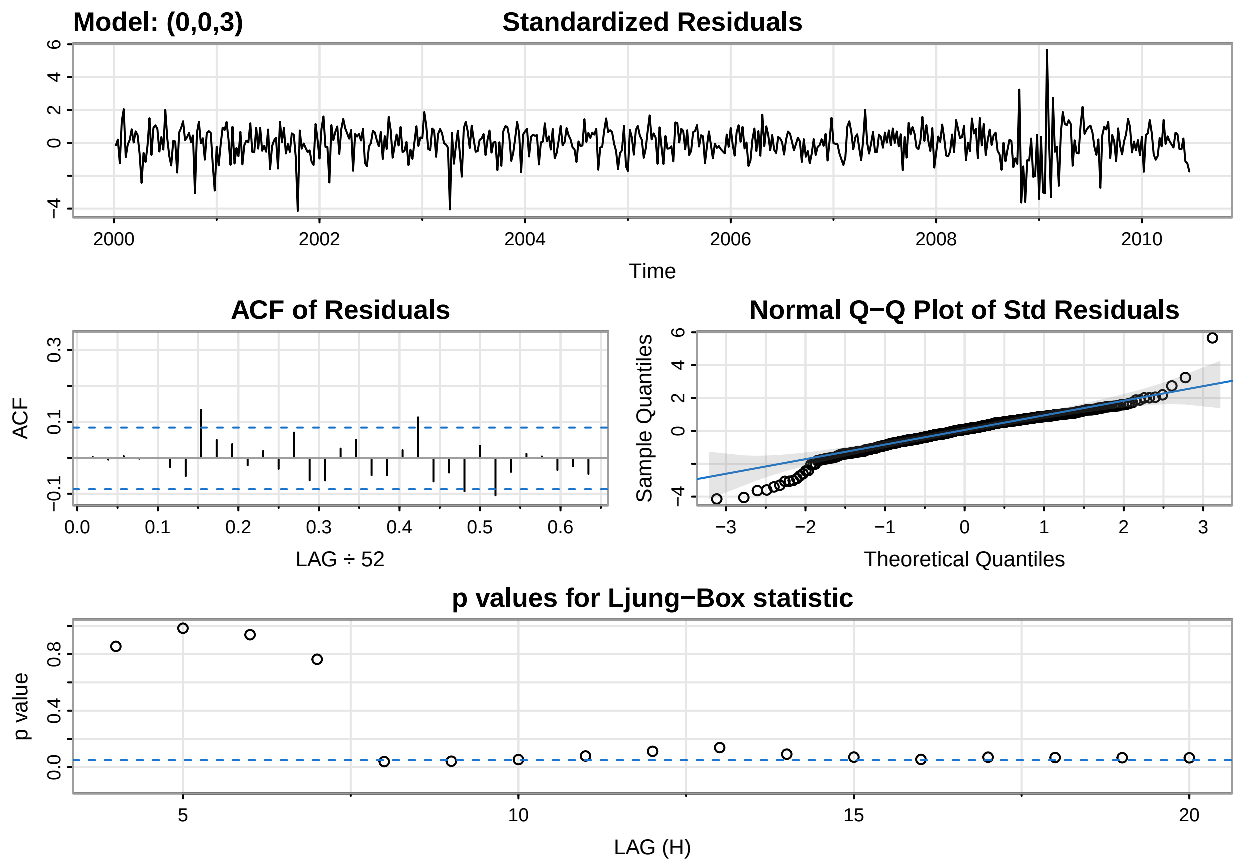


Figure 8: Diagnostics of the residuals from ARIMA(0,0,1) fit on crude oil prices per barrel growth rate.

Overall, ignoring the potential effect of the large outliers observed, the diagnostics have shown that the ARIMA(1,1,1) and ARIMA(0,1,3) models to the log of the oil prices or the ARIMA(1,0,1) and ARIMA(0,0,3) models to the growth rate appears to fit the data reasonably well.

To choose which is better, we compared the AIC, the AICc, and the BIC for the two models and the results are presented in Table 3. In terms of AIC and AICs the ARIMA(0,0,3) appears better for the growth rate, while ARIMA(1,0,1) appears better when considering the BIC.

	AIC	AICs	BIC
ARIMA(1,0,1)	-3.3121	-3.3120	-3.2805
ARIMA(0,0,3)	-3.3186	-3.3185	-3.2791

Table 3: Performance metrics for ARIMA(1,0,1) and ARIMA(0,0,3) for the growth rate of crude oil prices.

William Ofosu Agyapong
STAT 6391

Homework 7

Problem 4.10

(i)

First, we wish to show that if an infinite history, $\{x_n, x_{n-1}, \cdots, x_1, x_0, x_{-1}, \cdots\}$, is available, then

$$x^n_{n+m} = \sum_{j=0}^{\infty} \psi_j w^n_{m+n-j} = \sum_{j=m}^{\infty} \psi_j w_{m+n-j}.$$

Proof

For a future observation, x_{n+m}, we can write its causal representation as

$$x_{n+m} = \sum_{j=0}^{\infty} \psi_j w_{m+n-j} \qquad \text{for } m = 1, 2, \cdots$$

Then, the forecast for x_{n+m}, denoted x^n_{n+m}, is given by

$$x^n_{n+m} = \sum_{j=0}^{\infty} \psi_j w^n_{m+n-j}$$

Note that

$$w^n_{m+n-j} = E[w_{m+n-j} \mid w_1, w_2, \cdots, w_n] = w_{m+n-j}$$

And given the infinite history, $\{x_n, x_{n-1}, \cdots, x_1, x_0, x_{-1}, \cdots\}$, $x_0, x_{-1}, \cdots$ are usually not observed, so it is the case that

$$w^n_{m+n-j} = \begin{cases} w_{m+n-j}, & j \geq m \\ 0 & \text{otherwise} \end{cases}$$

Implying that

$$\begin{aligned} x^n_{n+m} &= \sum_{j=0}^{\infty} \psi_j w^n_{m+n-j} \\ &= \sum_{j=m}^{\infty} \psi_j w_{m+n-j}, \quad \text{as needed.} \quad \blacksquare \end{aligned}$$

Now, we use the result we just established to show that

$$E\left[x_{n+m} - x_{n+m}^{n}\right]^2 = E\left[\sum_{j=0}^{m-1} \psi_j w_{m+n-j}\right]^2 = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2$$

Proof

We have

$$E\left[x_{n+m} - x_{n+m}^{n}\right]^2 = E\left[\sum_{j=0}^{\infty} \psi_j w_{m+n-j} - \sum_{j=0}^{\infty} \psi_j w_{m+n-j}^{n}\right]^2$$

$$= E\left[\sum_{j=0}^{\infty} \psi_j w_{m+n-j} - \sum_{j=m}^{\infty} \psi_j w_{m+n-j}\right]^2 \quad \text{—— ②}$$

Since $j = 0, 1, 2, \cdots$ and $m = 1, 2, \cdots$, for any given value of m, terms in the summations for $j \geq m$ do not contribute anything to the expectation (their effects cancel out). To see this, we rewrite ② as follows

$$E\left[x_{n+m} - x_{n+m}^{n}\right]^2 = E\left[\left(\sum_{j=0}^{m-1} \psi_j w_{m+n-j} + \sum_{j=m}^{\infty} \psi_j w_{m+n-j}\right) - \sum_{j=m}^{\infty} \psi_j w_{m+n-j}\right]^2.$$

$$\Rightarrow \quad E\left[x_{n+m} - x_{n+m}^{n}\right]^2 = E\left[\sum_{j=0}^{m-1} \psi_j w_{m+n-j}\right]^2$$

$$= E\left[\sum_{j=0}^{m-1} \psi_j w_{m+n-j} \sum_{j=0}^{m-1} \psi_j w_{m+n-j}\right]$$

$$= E\left[\sum_{j=0}^{m-1} \sum_{k=0}^{m-1} \psi_j \psi_k w_{m+n-j} w_{m+n-k}\right]$$

$$= \sum_{j=0}^{m-1} \sum_{k=0}^{m-1} \psi_j \psi_k E\left(w_{m+n-j} w_{m+n-k}\right)$$

Noting that $E(w_t w_i) = 0$ for $t \neq i$ and $E(w_t w_i) = E(w_t^2) = \sigma_w^2$ for $t = i$, we have

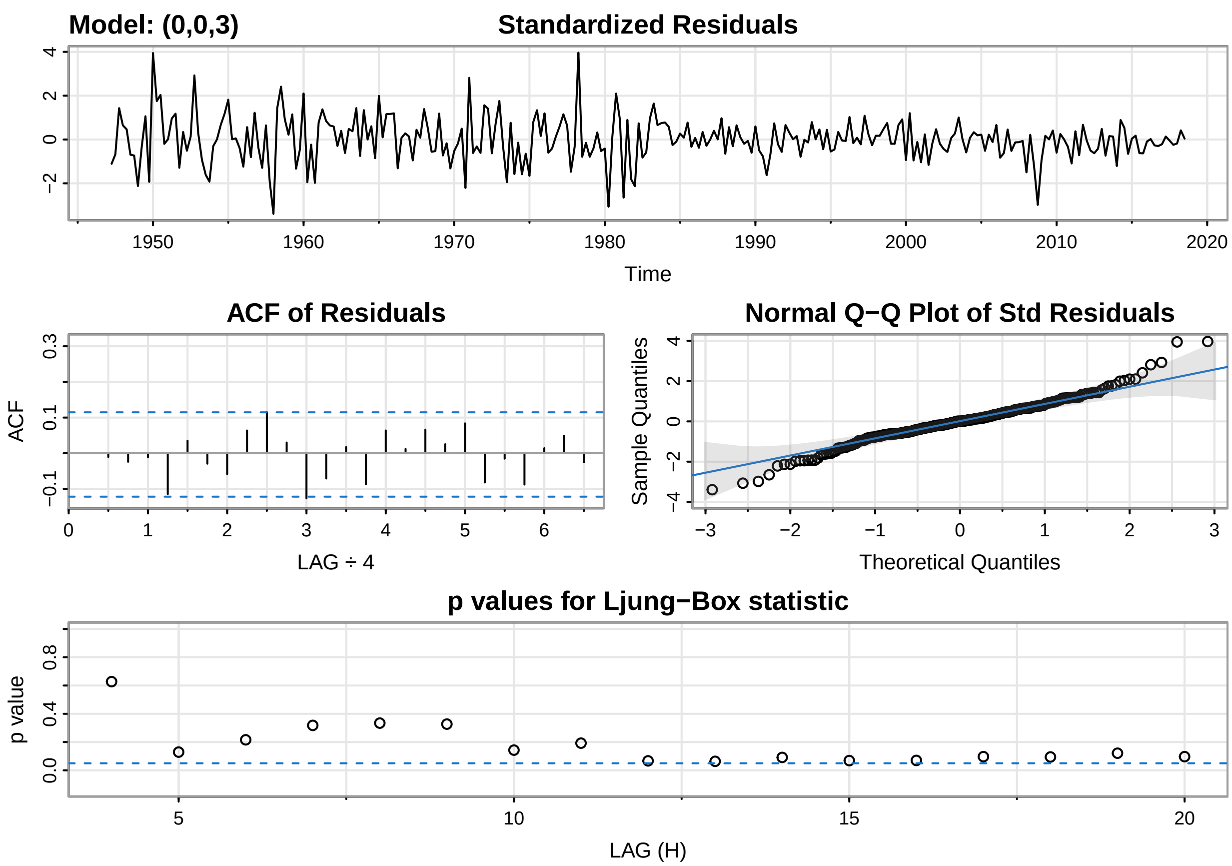
$$E\left[x_{n+m} - x_{n+m}^{n}\right]^2 = \sum_{j=0}^{m-1} \psi_j \psi_j \sigma_w^2, \quad \text{for } j = k$$

$$= \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2, \quad \text{as required.} \quad ▨$$

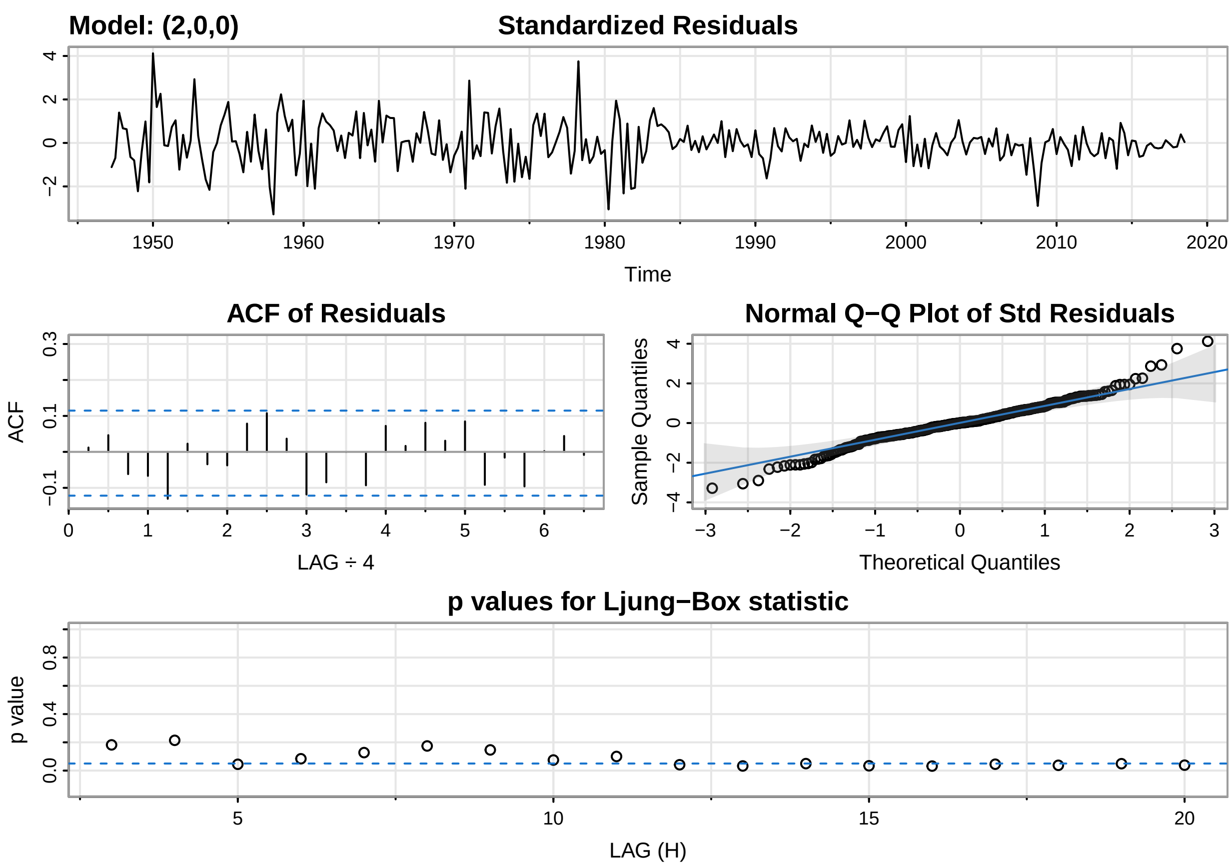
Appendix

A: Other models with extra parameters fitted to the GDP growth rate series

```
initial  value -4.672758
iter   2 value -4.750701
iter   3 value -4.754397
iter   4 value -4.754578
iter   5 value -4.754648
iter   6 value -4.754651
iter   7 value -4.754652
iter   8 value -4.754653
iter   9 value -4.754653
iter   9 value -4.754653
iter   9 value -4.754653
final  value -4.754653
converged
initial  value -4.755121
iter   2 value -4.755125
iter   3 value -4.755126
iter   4 value -4.755127
iter   5 value -4.755127
iter   6 value -4.755127
iter   6 value -4.755127
iter   6 value -4.755127
final  value -4.755127
converged
```



```
initial  value -4.673371
iter   2 value -4.737640
iter   3 value -4.748244
iter   4 value -4.748478
iter   5 value -4.748489
iter   6 value -4.748491
iter   7 value -4.748495
iter   8 value -4.748496
iter   8 value -4.748496
final  value -4.748496
converged
initial  value -4.748609
iter   2 value -4.748625
iter   3 value -4.748644
iter   4 value -4.748652
iter   5 value -4.748657
iter   6 value -4.748657
iter   6 value -4.748657
iter   6 value -4.748657
final  value -4.748657
converged
```



B: R Codes for the Analysis

```
# Set global options for output rendering
knitr::opts_chunk$set(eval = T, echo = F, warning = F, message = F,
```

```
                      fig.pos = "H", out.extra = "", fig.align = "center",
                      cache = F, comment="")

#----------------- Load required packages
# library(dplyr)
library(knitr)
library(kableExtra)
# library(broom)
library(stats)
library(astsa)


#----------------- set the current working directory to the file path
setwd(dirname(rstudioapi::getSourceEditorContext()$path))


set.seed(123)
lambda <- c(.25, .50, .75)
alpha <- 1 - lambda
x <- log(varve[1:100])
for (alpha_val in alpha) {
    ewma_log_varve <- HoltWinters(x, alpha = alpha_val, beta = FALSE, gamma = FALSE)
plot(ewma_log_varve, main="")
    # plot(x, type = "o", ylab = "log(varve)")
    # lines(ewma_log_varve$fit[,1], col=2)
}
#---------- Problem 5.2 codes

##-- Figure 2 --##
layout(1:2, heights=2:1)
tsplot(gdp, col=4)
acf_result <- acf1(gdp, main="")


##-- Figure 3 --##
tsplot(diff(log(gdp)), ylab="GDP Growth Rate", col=4)
abline(h=mean(diff(log(gdp))), col=6)

# plots the corresponding sample ACF and PACF
gdp_sample_acf <- acf1(diff(log(gdp)), main="")
gdp_sample_pacf <- acf1(diff(log(gdp)), pacf = T, main="")
ma2 <- sarima(diff(log(gdp)), 0,0,2, details = F) # MA(2) on growth rate

# collect relevant model outputs into a table
result_tbl <- data.frame(rbind(
    ma2$ttable,
    c(ma2$fit$sigma2, rep(NA, 3)),
    c(ma2$AIC, rep(NA, 3)),
    c(ma2$AICc, rep(NA, 3)),
    c(ma2$BIC, rep(NA, 3))
))
rownames(result_tbl) <- c("MA1", "MA2", "Constant", "Sigma^2", "AIC", "AICc", "BIC")
```

```
## display table of results
options(knitr.kable.NA='')  # suppress NAs from table output
result_tbl |>
    kable(booktabs=T, linesep="", align = "lcccc", digits=4,
          caption = "Parameter estimates and performance metrics for MA(2)  model for the GDP growth ra
    pack_rows("Coefficients", 1, 3) |>
    pack_rows("Performance metrics", 4, 7) |>
    kable_styling(latex_options = c("HOLD_position")) |>
    kable_classic()

#----- AR(1) estimation
ar1 <- sarima(diff(log(gdp)), 1,0,0, details = F) # AR(1) on growth rate

# collect relevant model outputs into a table
result_tbl <- data.frame(rbind(
    ar1$ttable,
    c(ar1$fit$sigma2, rep(NA, 3)),
    c(ar1$AIC, rep(NA, 3)),
    c(ar1$AICc, rep(NA, 3)),
    c(ar1$BIC, rep(NA, 3))
))
rownames(result_tbl) <- c("AR1", "Constant", "Sigma^2", "AIC", "AICc", "BIC")

## display table of results
options(knitr.kable.NA='')  # suppress NAs from table output
result_tbl |>
    kable(booktabs=T, linesep="", align = "lcccc", digits=4,
          caption = "Parameter estimates and performance metrics for AR(1)  model for the GDP growth ra
    pack_rows("Coefficients", 1, 2) |>
    pack_rows("Performance metrics", 3, 6) |>
    kable_styling(latex_options = c("HOLD_position")) |>
    kable_classic()


ma2 <- sarima(diff(log(gdp)), 0,0,2) # MA(2) on growth rate


ar1 <- sarima(diff(log(gdp)), 1,0,0) # AR(1) on growth rate

##--------- Problem 5.3 codes

# plot the original series data
tsplot(oil, col=4, ylab = "Prices in dollars per barrel")

# plot of the growth rate
oil_growth_rate <- diff(log(oil))
tsplot(oil_growth_rate, ylab = "growth rate", col = 4)
sample_acf <- acf1(oil_growth_rate)
sample_pacf <- acf1(oil_growth_rate, pacf = T)

arma11_fit <- sarima(oil_growth_rate, 1,0,1) # ARMA(1,1) to the growth rate
arma03_fit <- sarima(oil_growth_rate, 0,0,3) # ARMA(0,3) to the growth rate
# collect relevant model outputs into a table
```

```r
result_tbl <- data.frame(rbind(

    c(arma11_fit$AIC, arma11_fit$AICc, arma11_fit$BIC),
    c(arma03_fit$AIC, arma03_fit$AICc, arma03_fit$BIC)
))
names(result_tbl) <- c("AIC", "AICs", "BIC")
rownames(result_tbl) <- c("ARIMA(1,0,1)", "ARIMA(0,0,3)")

## display table of results
result_tbl |>
    kable(booktabs=T, linesep="", align = "ccc", digits=4,
          caption = "Performance metrics for ARIMA(1,0,1) and ARIMA(0,0,3) for the growth rate of crude
    kable_styling(latex_options = c("HOLD_position")) |>
    kable_classic()
# trying other forms of models:
other_mod<- sarima(diff(log(gdp)), 0, 0, 3) # try an MA(2+1) fit (not shown)
other_mod2 <- sarima(diff(log(gdp)), 2, 0, 0) # try an AR(1+1) fit (not shown)

# sarima(diff(log(gdp)), 0, 0, 4) # try an MA(2+2) fit (not shown)
# sarima(diff(log(gdp)), 3, 0, 0) # try an AR(1+2) fit (not shown)
```

References

Shumway, Robert, and David Stoffer. 2019. *Time Series: A Data Analysis Approach Using r*. CRC Press.