Homework 1: Time Series Elements, Correlation and Stationary Time Series

Time Series Analysis (STAT 6391)

Willliam Ofosu Agyapong*

September 07, 2023

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

1 Problem 1.1

1.1 Part (a)

The plot below depicts 100 observations from the autogression

$$x_t = -.9x_{t-2} + w_t.$$

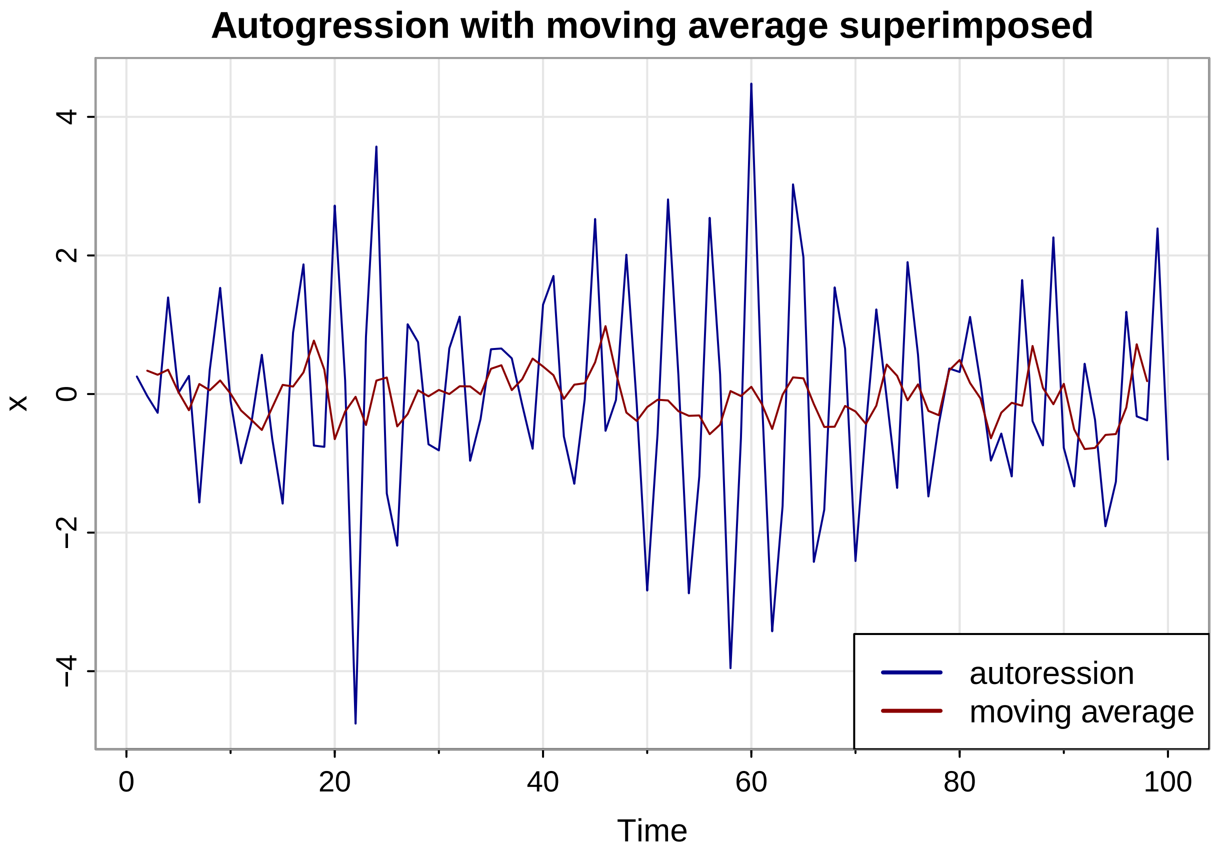

Figure 1: Autogression with moving average superimposed

1.2 Part (b)

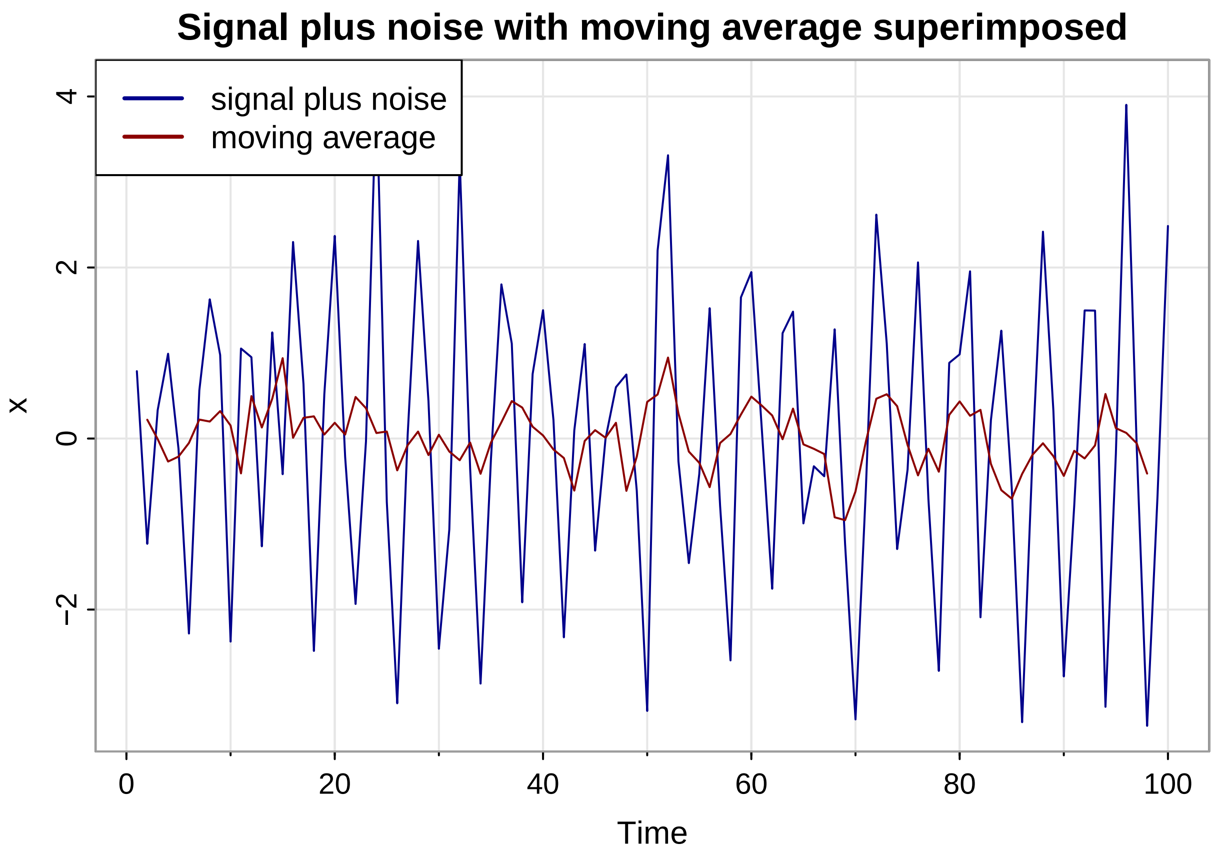

Figure 2: Signal plus noise with moving average superimposed

1.3 Part (c)

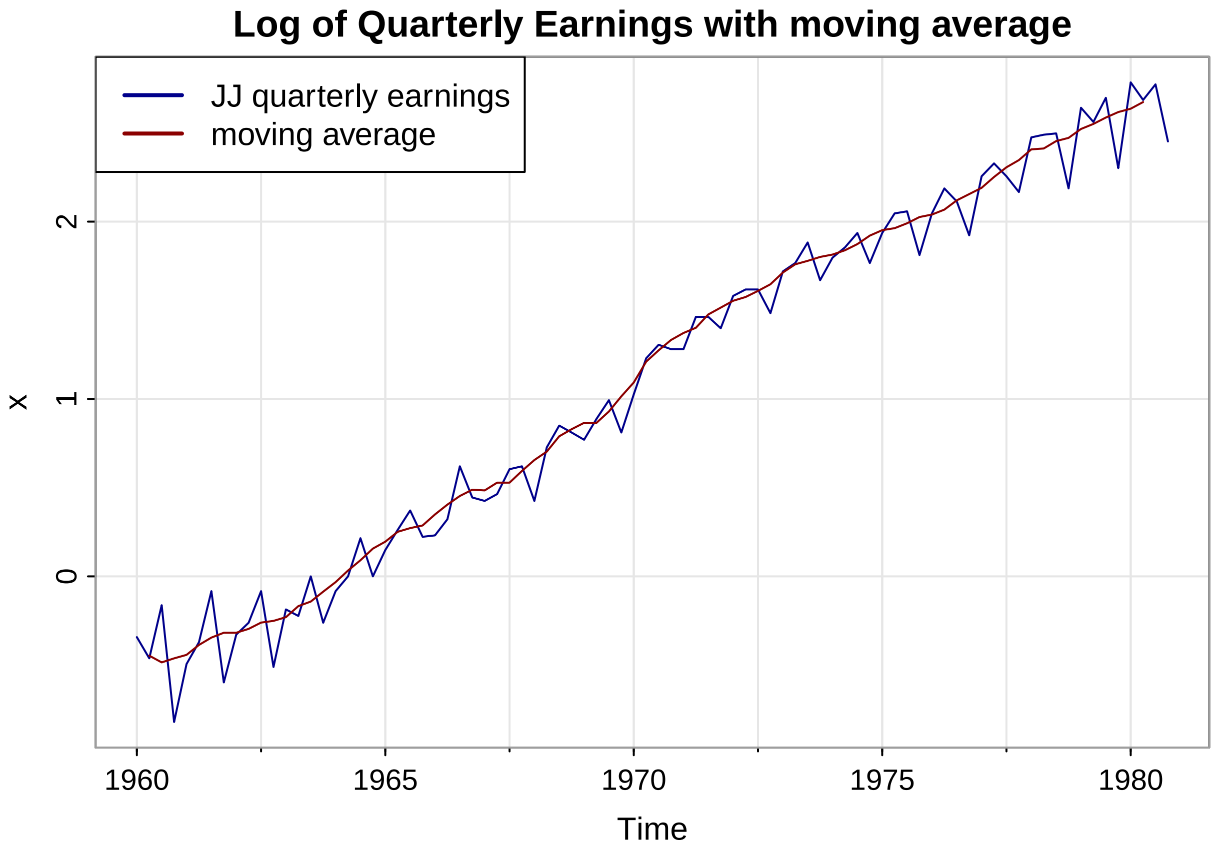

Figure 3: Log of Johson & Johnson Quarterly Earnings with moving average superimposed

1.4 Part (d): What is seasonal adjustment?

Seasonal adjustment is a statistical technique used to remove the seasonal component of a time series in order to better understand its underlying trend and cyclical patterns, which allows for more meaningful comparisons of economic conditions from period to period. This process usually involves decomposing a time series into four separate components: (1) the trend-cycle, (2) seasonal effects, (3) other calendar effects such as trading days and moving holidays, and (4) the irregular component. The seasonally adjusted series is the original time series with the estimated seasonal and calendar effects removed. The seasonally adjusted data Smoothing techniques such as the moving average is a useful method for discovering seasonal components.

1.5 Part (e): Conclusion

The main take away for me from this exercise is that applying moving average to the series smoothed the data, leading to much reduction in variance and removal of fast oscillations.

2 Problem 1.2

2.1 Part (a)

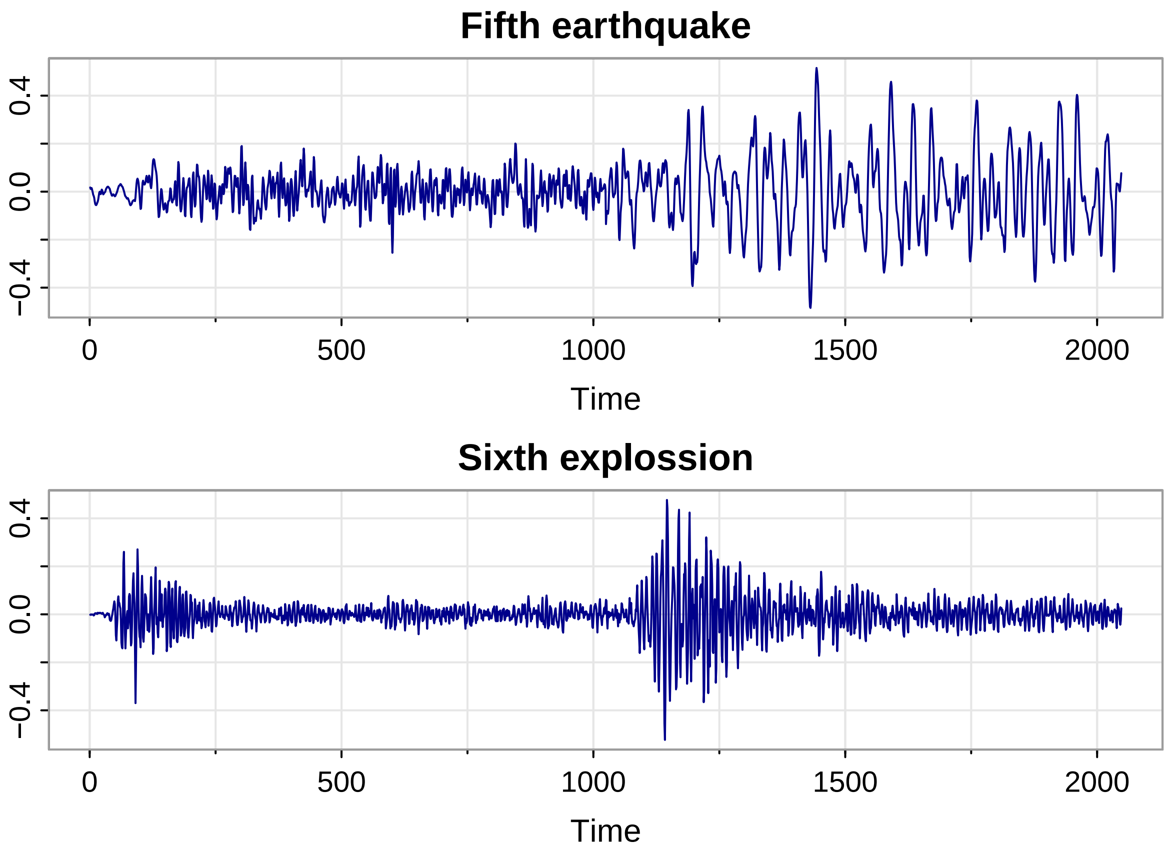

Figure 4: Seismic recordings

2.2 Part (b)

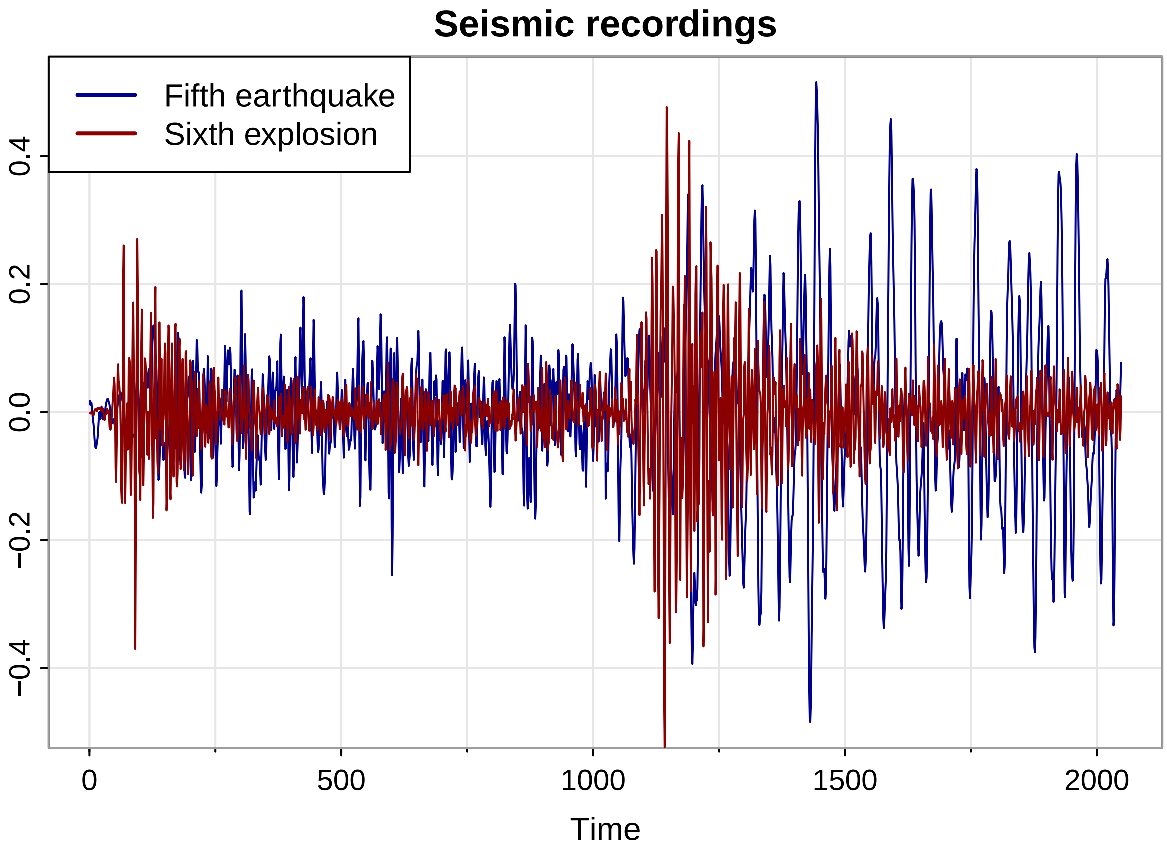

Figure 5: Comparing seismic recordings of earthquake and explossions

2.3 Part (c): In what way are the earthquake and explosion series different?

We can see from the plots that, the explosion time series have sharp, high-frequency signals that decay rapidly, while the earthquake series exhibits low-frequency signals that last longer (much slower decay).

3 Problem 1.3

3.1 Part (a)

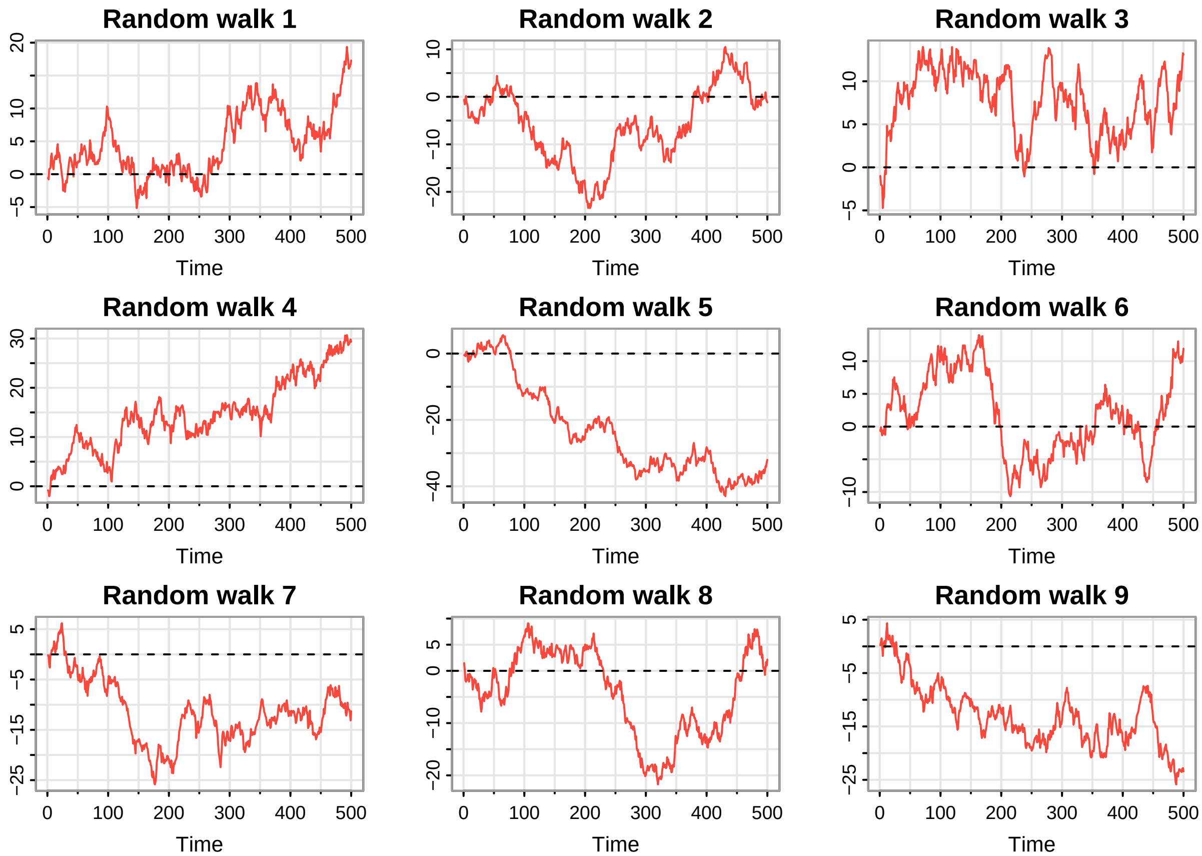

Figure 6: Nine series generated from random walks without drift

3.2 Part (b)

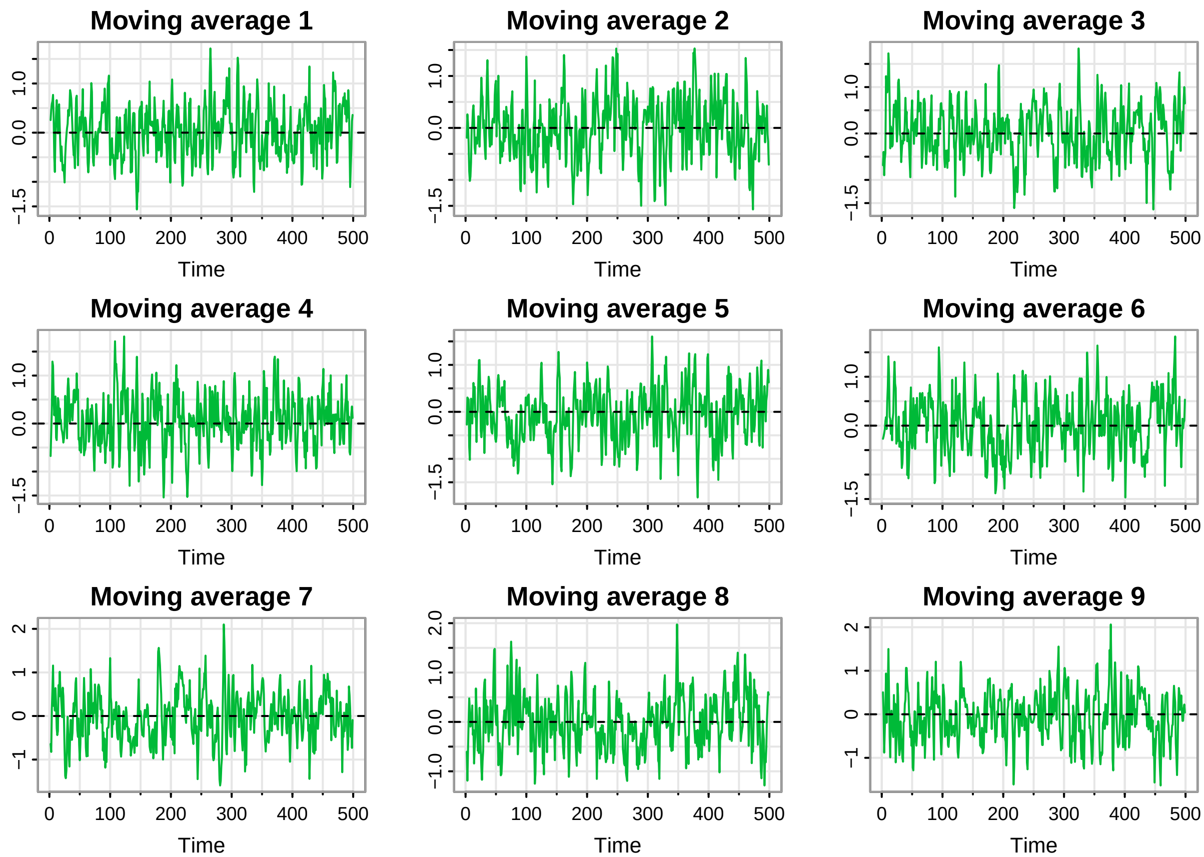

Figure 7: Nine moving average series

3.3 Part (c)

The main differences between the results of part (a) and part (b) are summarized as follows:

The random walk series in part (a) are too erratic (more variable) to the extent that one cannot observe the same behavior or pattern in the nine series apart from the fact that they almost always start at zero and assume highly unpredictable values thereafter over time. On the other hand, the moving average series in part (b) exhibit the opposite characteristics; they are all centered around zero (0) and remained quite stable over time with relatively low variability (all are within reasonable range).

4 Problem 1.4

4.1 Part (a)

The plot of seasonally adjusted GDP in the figure below can be compared to the random ***walk with drift*** model as depicted in Figure 1.10 of Section 1.3 in (Shumway and Stoffe, 2019). We notice that, while both exhibit an upward trend which is not linear, the plot here is much smoother.

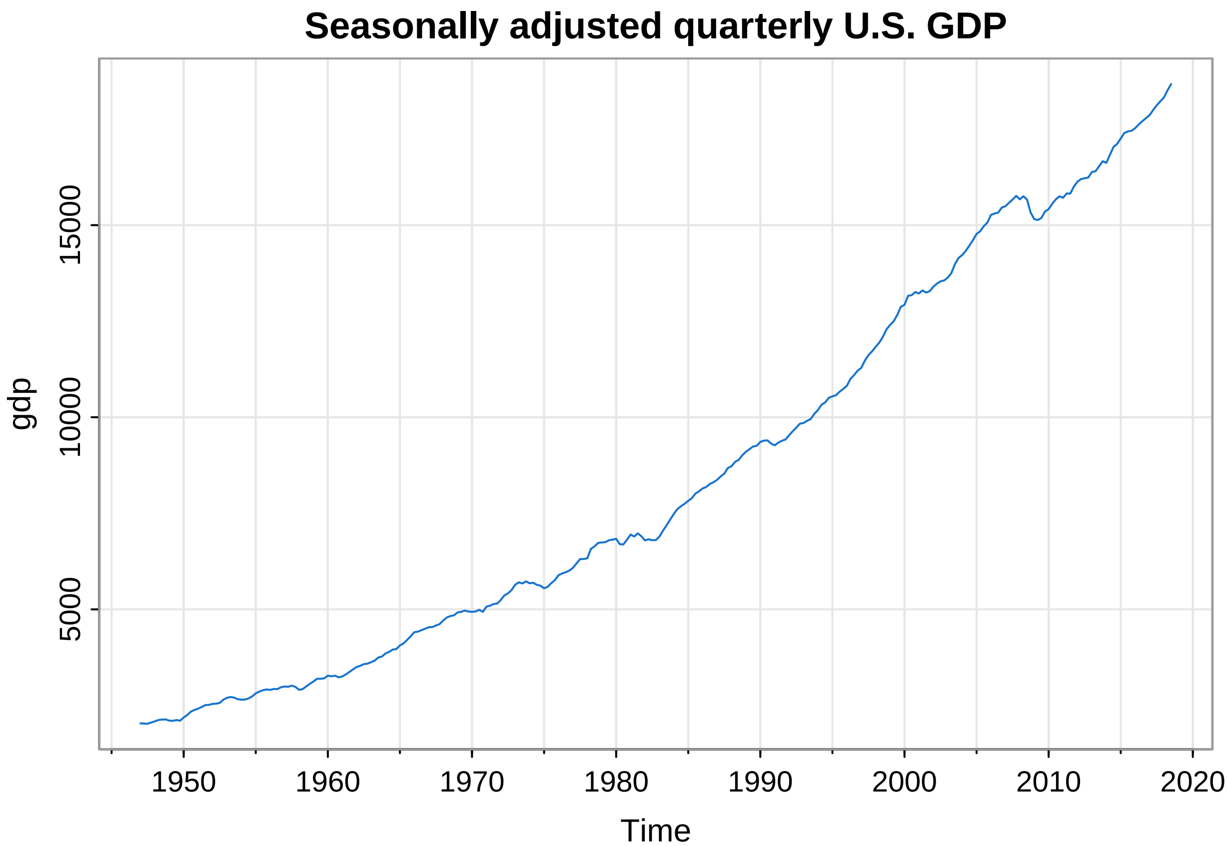

Figure 8: Seasonally adjusted quarterly U.S. GDP from 1947-I to 2018-III.

4.2 Part (b)

Figure 1.4 of Shumway and Stoffe (2019) is reproduced as follows. Interestingly, the two methods of calculating growth rate produce identical if not the same results as we see the square characters and the x characters overlapping for almost all the time points. Roughly speaking, there is no difference between the end results of the two methods.

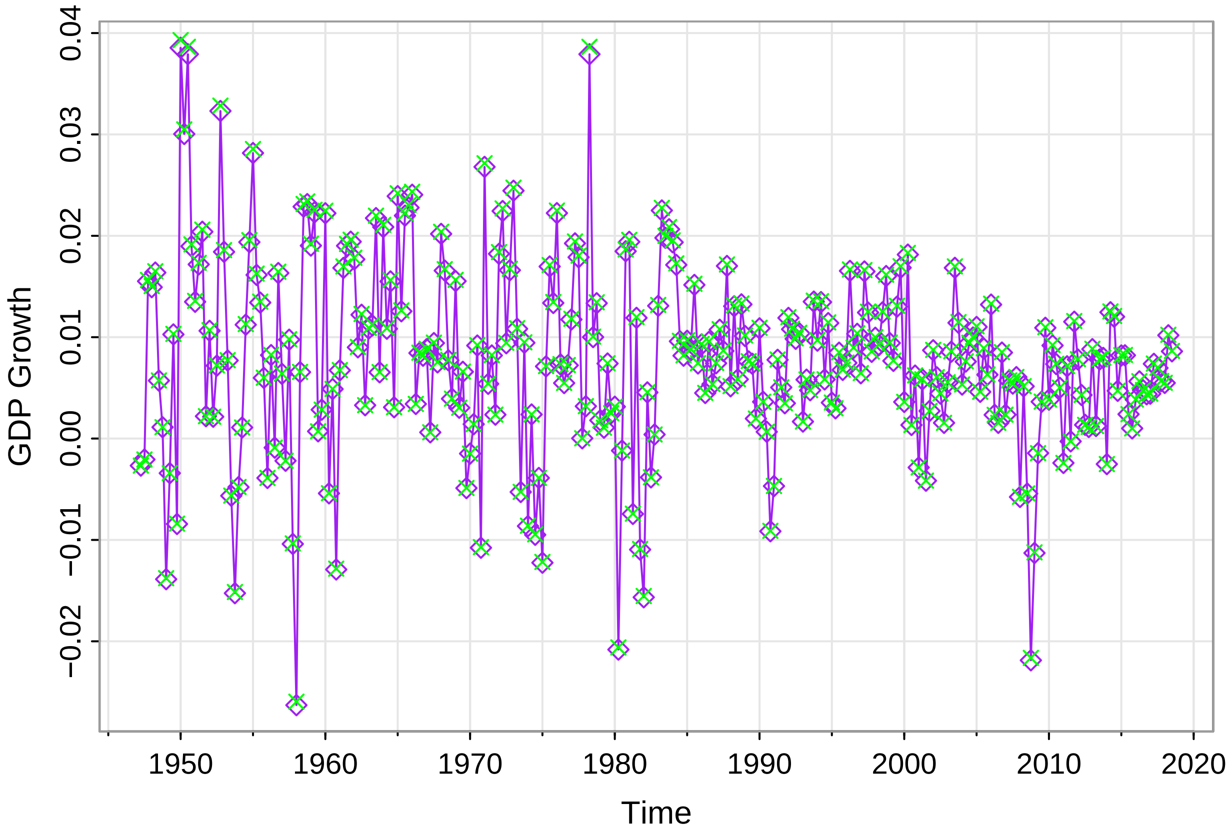

Figure 9: US GDP growth rate calculated using logs (purple square) and actual values (green x)

4.3 Part (c)

From the above figure, it is clear that the behavior of the growth in U.S GDP can best be modeled by an ***autoregression*** model.

5 Problem 2.2

Please see solution in the attachment.

Homework 1 (Problem 2.2)

Consider the time series
$$x_t = \beta_0 + \beta_1 t + w_t,$$
where β_0 and β_1 are regression coefficients, and w_t is a white noise process with variance σ_w^2.

(a) We wish to determine whether x_t is stationary, which calls for the derivation of the mean and covariance functions.

For the mean function, we have

$$\mu_t = E(\beta_0 + \beta_1 t + w_t) = \beta_0 + \beta_1 t + E(w_t)$$
$$= \beta_0 + \beta_1 t + 0 = \beta_0 + \beta_1 t,$$

which depends on the time, t, and for this reason x_t cannot be stationary.
The covariance function is also given as

$$\gamma_x(s,t) = Cov(x_s, x_t) = Cov[(\beta_0 + \beta_1 s + w_s), (\beta_0 + \beta_1 t + w_t)]$$

when $s = t$
$$\Rightarrow \gamma_x(t,t) = Cov[(\beta_0 + \beta_1 t + w_t), (\beta_0 + \beta_1 t + w_t)]$$
$$= Cov(w_t, w_t) = Var(w_t) = \sigma_w^2$$

when $s = t+1$
$$\Rightarrow \gamma_x(t+1,t) = Cov[(\beta_0 + \beta_1(t+1) + w_{t+1}), (\beta_0 + \beta_1 t + w_t)]$$
$$= 0$$

Generally,
$$\gamma_x(s,t) = \begin{cases} \sigma_w^2 & s = t \text{ at lag } 0 \\ 0 & |s-t| \geq 1 \end{cases}$$
depends on s and t only through the lag.

Hence, we conclude that x_t is not stationary since its mean function depends on t. However, it is evident that there is trend stationarity since the autocovariance function only depends on the lag.

Part (b)

Show that the process $y_t = x_t - x_{t-1}$ is stationary.

We have

$$y_t = x_t - x_{t-1} = (\beta_0 + \beta_1 t + w_t) - (\beta_0 + \beta_1(t-1) + w_{t-1})$$
$$= \beta_0 + \beta_1 t + w_t - \beta_0 - \beta_1 t + \beta_1 - w_{t-1}$$
$$= \beta_1 - w_{t-1}$$

$\Rightarrow \mu_{y,t} = E[y_t] = \beta_1 - E(w_{t-1}) = \beta_1$, independent of t.

$$\gamma_y(h) = Cov(y_{t+h}, y_t) = E[(y_{t+h} - \mu_{y,t+h})(y_t - \mu_{y,t})]$$
$$= E\{[\beta_1 + w_{t+h} - w_{t+h-1} - \beta_1][\beta_1 + w_t - w_{t-1} - \beta_1]\}$$
$$= E[(w_{t+h} - w_{t+h-1})(w_t - w_{t-1})]$$

when $h = 0$

$$\gamma_y(0) = E[(w_t - w_{t-1})(w_t - w_{t-1})] = Var(w_t) + Var(w_{t-1})$$
$$= 2\sigma_w^2$$

when $h = 1$

$$\gamma_y(1) = E[(w_{t+1} - w_t)(w_t - w_{t-1})] = -E(w_t \cdot w_t) = -\sigma_w^2$$
$$= \gamma_y(-1)$$

when $h = 2$

$$\gamma_y(2) = E[(w_{t+2} - w_{t+1})(w_t - w_{t-1})] = 0 = \gamma_y(-2).$$

Thus

$$\gamma_y(h) = \begin{cases} 2\sigma_w^2 & h = 0 \\ -\sigma_w^2 & |h| = 1 \\ 0 & |h| \geq 1 \end{cases}$$

, which depends only on the lag, h.

Therefore,

it follows that the process y_t is stationary since its mean function, $\mu_{y,t}$ is a constant (β_1) which does not depend on t and its autocovariance function, $\gamma_y(h)$ only depends on the time difference h.

Part (c)

Show that the mean of the two-sided moving average

$$v_t = \frac{1}{3}(x_{t-1} + x_t + x_{t+1})$$

is $\beta_0 + \beta_1 t$.

$x_t = \beta_0 + \beta_1 t + w_t$

Proof:

$$\mu_{v,t} = E[v_t] = \frac{1}{3}\left[E(x_{t-1}) + E(x_t) + E(x_{t+1})\right]$$

$$= \frac{1}{3}\left\{[\beta_0 + \beta_1(t-1)] + (\beta_0 + \beta_1 t) + (\beta_0 + \beta_1(t+1))\right\}$$

$$= \frac{1}{3}\left[\beta_0 + \beta_1 t - \beta_1 + \beta_0 + \beta_1 t + \beta_0 + \beta_1 t + \beta_1\right]$$

$$= \frac{1}{3}\left[3\beta_0 + 3\beta_1 t\right] = \frac{1}{3}\left[3(\beta_0 + \beta_1 t)\right] = \beta_0 + \beta_1 t$$

$$= \beta_0 + \beta_1 t$$

as required.

Appendix

A: R codes

```
# 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)

#----------------- Load required packages
library(stats)
library(astsa)


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




#------ Problem 1.1

#-- generate 100 observations
set.seed(123) # seed for reproducible results

# first generate the white noise process with sigma = 1
wn <- rnorm(100 + 50)[-c(1:50)] # added extra 50 to avoid startup problems from initialization

# next, generate the autoregressive process
x <- stats::filter(wn, filter = c(0, -.9), method = "recursive")

# apply the moving average filter
v <- stats::filter(x,sides = 2, filter = rep(.25, 4), method = "convolution")

tsplot(x, col="darkblue", main = "Autogression with moving average superimposed")
lines(v, col="darkred")
legend("bottomright", legend = c("autoression", "moving average"),
       lty = 1, lwd = 2, bg="white",
       col = c("darkblue", "darkred"))


#-- generate 100 observations
t <- 1:100
signal <- 2*cos(2*pi*t/4)
wn <- rnorm(100) # white noise  with sigma = 1
# next, generate series
x <- signal + wn

# apply the moving average filter
v <- stats::filter(x, sides = 2, filter = rep(.25, 4), method = "convolution")

tsplot(x, col="darkblue", main = "Signal plus noise with moving average superimposed")
lines(v, col="darkred")
legend("topleft", legend = c("signal plus noise", "moving average"),
       lty = 1, lwd = 2, bg="white",
```

```
       col = c("darkblue", "darkred"))
#-- generate 100 observations

x <- log(astsa::jj)

# apply the moving average filter
v <- stats::filter(x, sides = 2, filter = rep(.25, 4), method = "convolution")

tsplot(x, col="darkblue", main = "Log of Quarterly Earnings with moving average")
lines(v, col="darkred")
legend("topleft", legend = c("JJ quarterly earnings", "moving average"),
       lty = 1, lwd = 2, bg="white",
       col = c("darkblue", "darkred"))

#---------- Problem 1.2

# Part (a)
par(mfrow=2:1)
# plot the fifth earthquake
tsplot(EQ5, main = "Fifth earthquake", col="darkblue", ylab = "")
# plot the sixth explossion
tsplot(EXP6, main = "Sixth explossion", col = "darkblue", ylab = "")


# Part (b)

tsplot(EQ5, main = "Seismic recordings", col="darkblue", ylab = "")
lines(EXP6, col="darkred")
legend("topleft", col = c("darkblue", "darkred"), lty = 1, lwd = 2, bg="white",
       legend = c("Fifth earthquake", "Sixth explosion"))




#---------- Problem 1.3

# Part (b)
set.seed(123)

par(mfrow=c(3,3))

for (i in 1:9) {
    wn <- rnorm(500); x <- cumsum(wn) # random walk without drift
    # wnd <- wn + 1; xd <- cumsum(wnd) # random walk with drift

    tsplot(x, main = paste("Random walk", i), col = 2, ylab = "")
    # abline(a=0, b=1, col=4, lty=2) # plot the drift
    # lines(x, col="darkred")
    abline(h=0, lty=2)
}


# Part (b)
```

```r
set.seed(123)

par(mfrow=c(3,3)) # for multiple plot

for (i in 1:9) {
    wn <- rnorm(500)
    v <- filter(wn, sides = 2, filter = rep(1/3, 3), method = "convolution")
    tsplot(v, main = paste("Moving average",i), col = 11, ylab = "")
    abline(h=0, lty=2)
}



#---------- Promblem 1.4

# Part (a)
tsplot(gdp, col = 4, main = "Seasonally adjusted quarterly U.S. GDP")


# Part (b)
tsplot(diff(log(gdp)), type = "o", pch=5, col = "purple", ylab = "GDP Growth")
points(diff(gdp)/lag(gdp, -1), pch=4, col="green")

# xfun::embed_file("problem 2.2.pdf")
```

References

- Robert H. Shumway, & David S. Stoffer. (2019). Time Series: A Data Analysis Approach Using R
- Seasonally adjusted data - Frequently asked questions (Statistics Canada). Accessed on 09/07/2023 at https://www150.statcan.gc.ca/n1/dai-quo/btd-add/btd-add-eng.htm#