



# ST565 Homework 2

Brian Cervantes Alvarez

February 19, 2024

## Problem 1

### Part A

Recall that a process is stationary if its mean and variance are constant over time, and its autocovariance function depends only on the lag between two time points, not on the actual time points themselves.

Given  $\{x_t\}$ , the mean function is  $\mu_t = \beta_1 + \beta_2 t$ , which is a linear function of time. Hence, the mean is not constant, and  $\{x_t\}$  is not stationary.

### Part B

We need to compute  $y_t$ :

$$y_t = (\beta_1 + \beta_2 t + w_t) - (\beta_1 + \beta_2(t-1) + w_{t-1}) = \beta_2 + w_t - w_{t-1}$$

Here, the mean of  $y_t$  is  $\beta_2$  is constant, and the process depends only on the white noise difference  $(w_t - w_{t-1})$ , suggesting  $y_t$  could be stationary. However, in order to do the full analysis we should consider the variance and autocovariance, which are expected to be constant and only dependent on the lag, respectively.

### Part C

We know that  $w_t$  are iid with variance  $\sigma^2$ , the autocovariance of  $v_t$  for lags  $|k| \leq q$  can be simplified to:

$$v_t = \frac{1}{2q+1} \sum_{r=-q}^q x_{t-r}$$

$$\gamma_v(k) = \text{Cov}(v_t, v_{t+k}) = \frac{1}{(2q+1)^2} \sum_{r=-q}^q \sum_{s=-q}^q \text{Cov}(w_{t-r}, w_{t+k-s})$$

$$\gamma_v(k) = \frac{\sigma^2}{(2q+1)^2} \times (2q+1 - |k|) = \frac{\sigma^2(2q+1 - |k|)}{(2q+1)^2}$$

For  $|k| > q$ , they do not overlap, and thus  $\gamma_v(k) = 0$ , Hence autocovariance of  $v_t$  depends on the lag  $k$



---

## Problem 2

---

### Part A

---

$$x_t = 0.3x_{t-1} + w_t \Rightarrow (1 - 0.3B)x_t = w_t$$

### Part B

---

$$x_t = w_t - 1.3w_{t-1} + 0.4w_{t-2} \Rightarrow x_t = (1 - 1.3B + 0.4B^2)w_t$$

### Part C

---

$$x_t = 0.5x_{t-1} + w_t - 1.3w_{t-1} + 0.4w_{t-2} \Rightarrow (1 - 0.5B)x_t = (1 - 1.3B + 0.4B^2)w_t$$

### Part D

---

- Model A is stationary if the root of  $1 - 0.3B = 0$  lies outside the unit circle, which it does ( $B \neq \frac{1}{0.3}$ ).
- Model B requires analyzing the roots of  $1 - 1.3B + 0.4B^2 = 0$ . If  $|B| > 1$ , then the process is stationary.



## Problem 3

### Part A

```
library(tidyverse)
library(forecast)

ts1 <- read_csv("ts1.csv") # Has one variable called x
ts2 <- read_csv("ts2.csv") # Has one variable called x

# Add an index column to act as time
ts1 <- ts1 %>% mutate(time = row_number())
ts2 <- ts2 %>% mutate(time = row_number())

# For ts1
model_ts1 <- lm(x ~ time, data = ts1)
detrended_ts1 <- residuals(model_ts1)

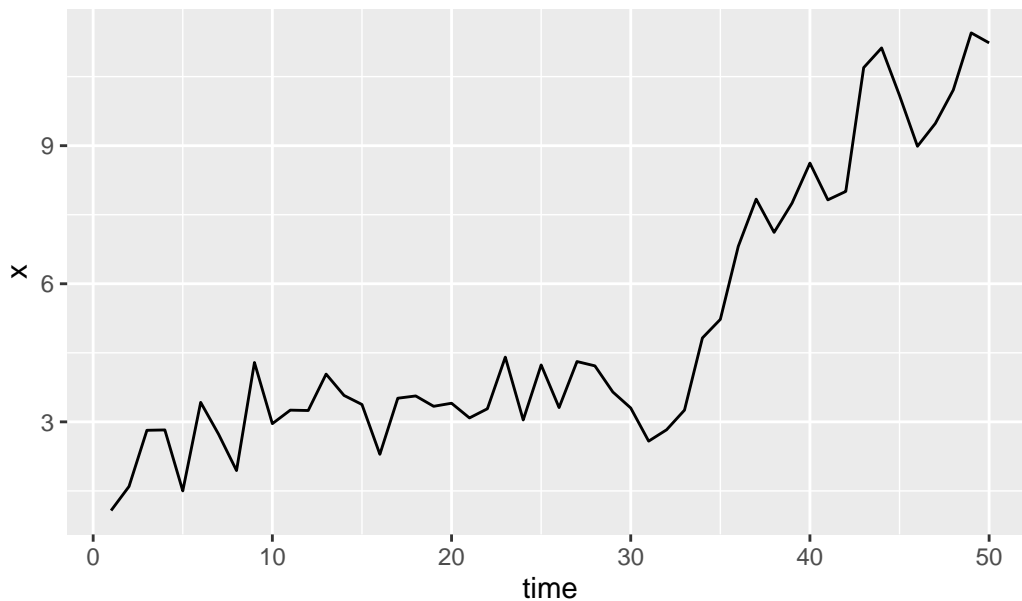
# For ts2
model_ts2 <- lm(x ~ time, data = ts2)
detrended_ts2 <- residuals(model_ts2)
```

## Part B

```
diff_ts1 <- diff(ts1$x)
diff_ts2 <- diff(ts2$x)

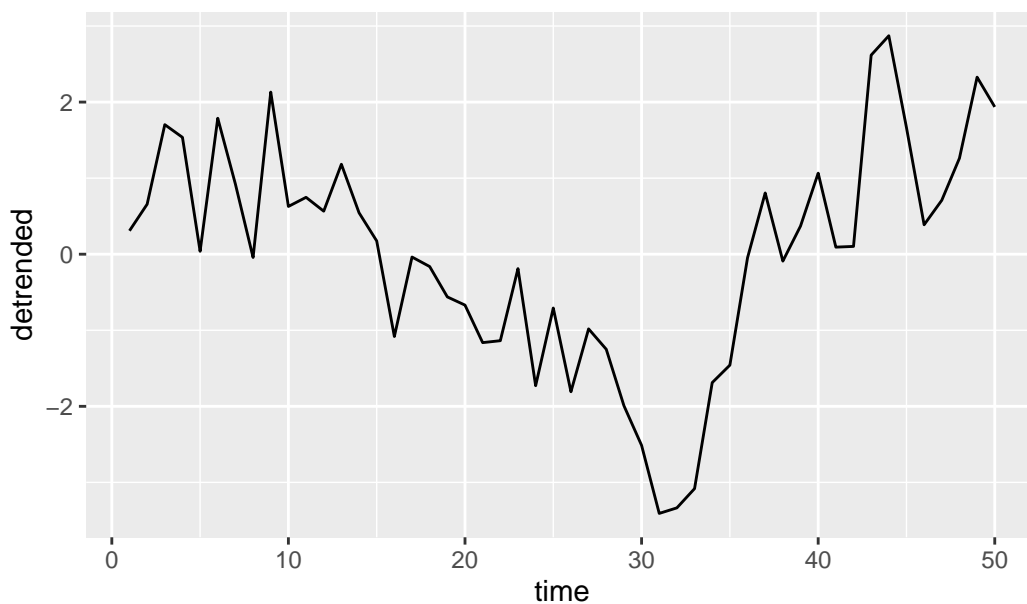
# Original Series
ggplot(ts1, aes(x = time, y = x)) + geom_line() + ggtitle("Original Series")
```

Original Series



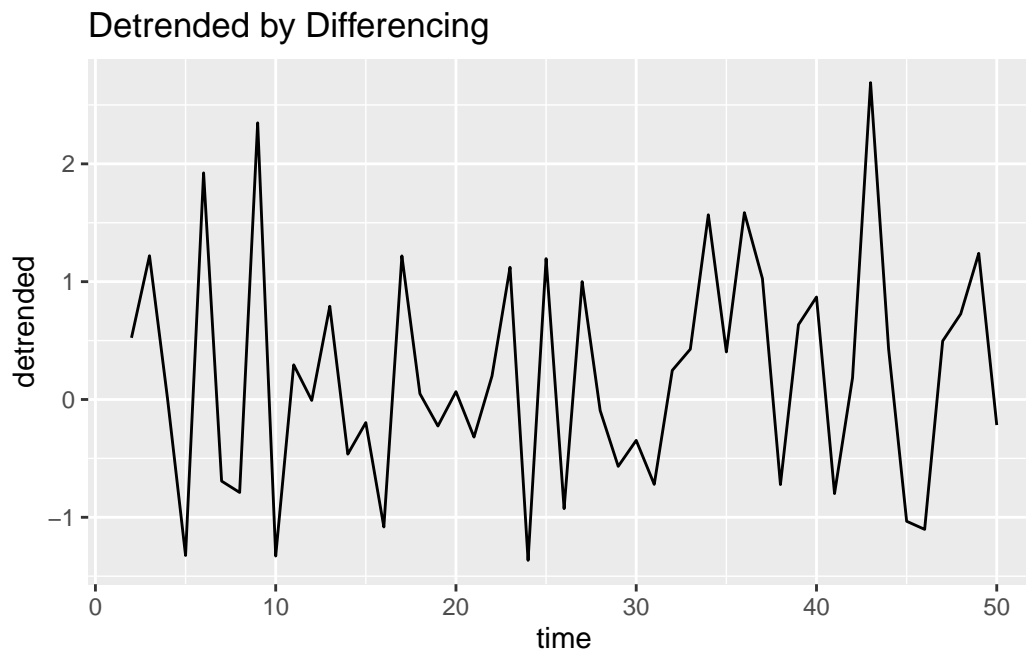
```
# Detrended by Regression
ggplot(data.frame(time = ts1$time, detrended = detrended_ts1), aes(x = time, y = detrended)) +
  geom_line() + ggtitle("Detrended by Regression")
```

Detrended by Regression



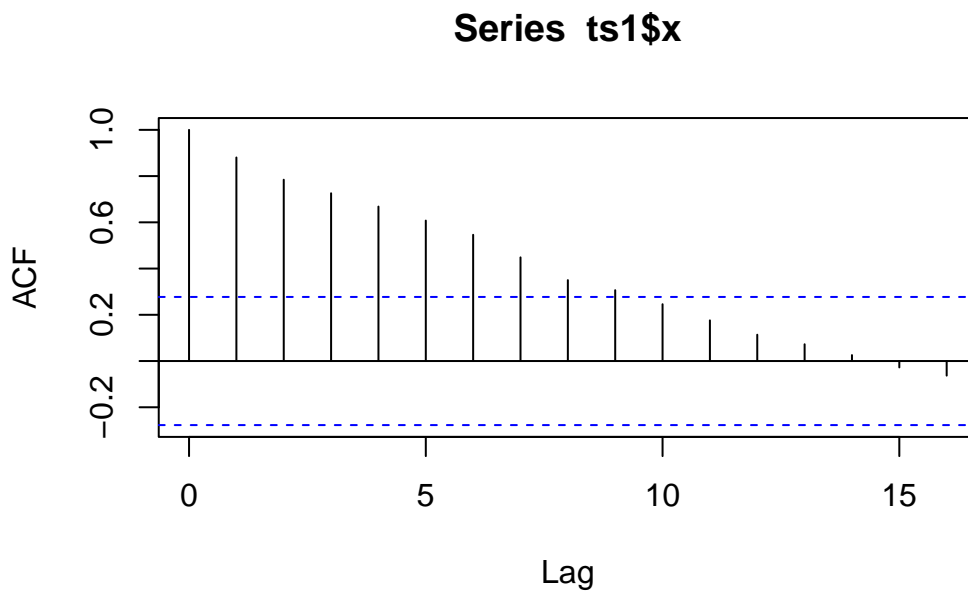


```
# Detrended by Differencing  
ggplot(data.frame(time = ts1$time[-1], detrended = diff_ts1), aes(x = time, y = detrended))  
  geom_line() + ggtitle("Detrended by Differencing")
```

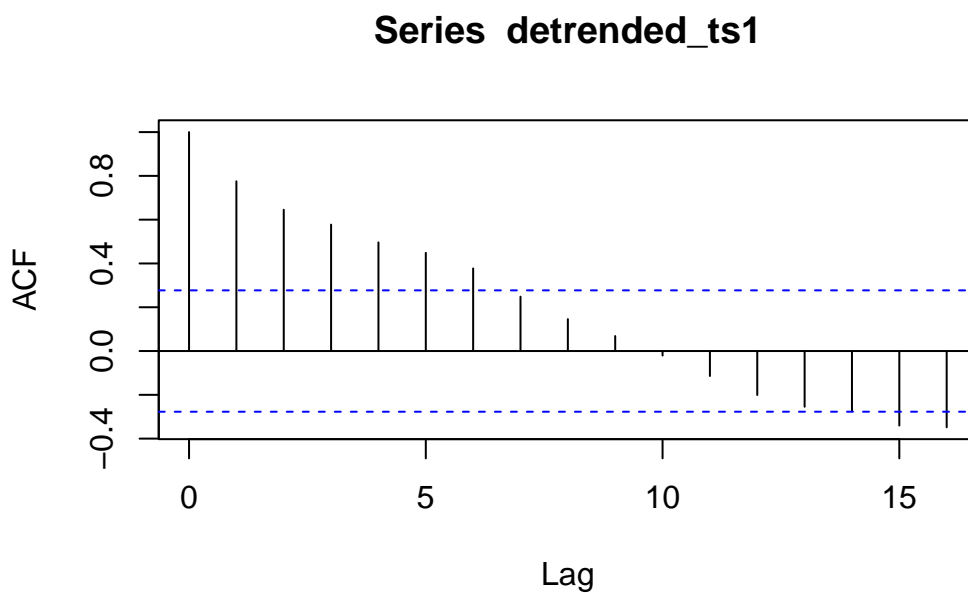


## Part C

```
# 1st Time Series  
# ACF of ts1  
acf(ts1$x)
```

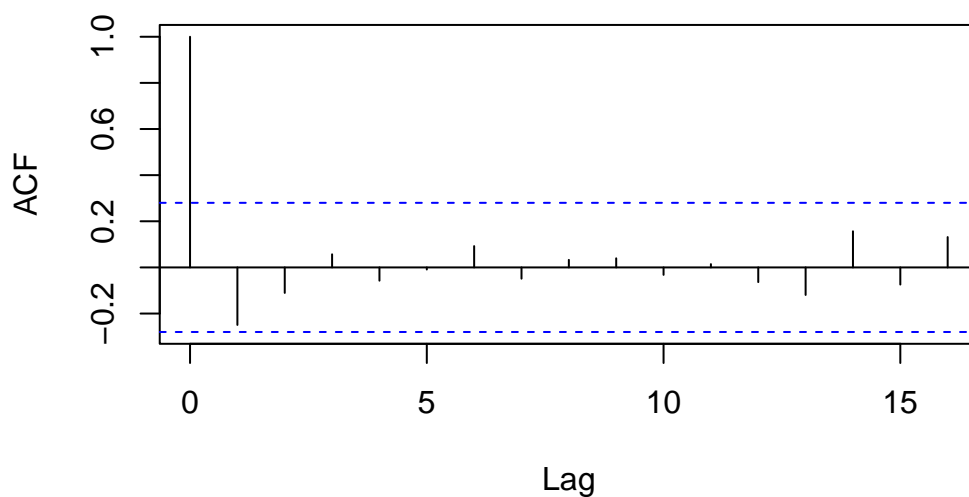


```
# ACF of Detrended by Regression  
acf(detrended_ts1)
```



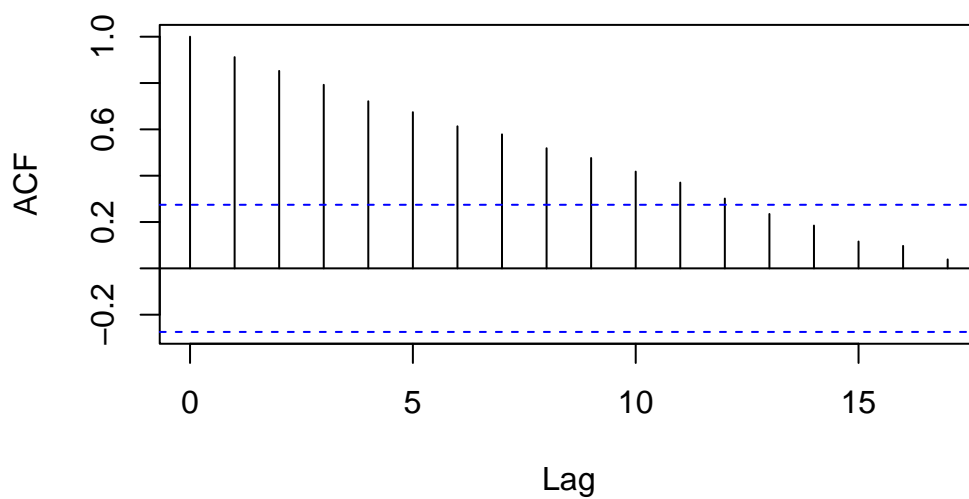
```
# ACF of Detrended by Differencing  
acf(diff_ts1)
```

### Series diff\_ts1



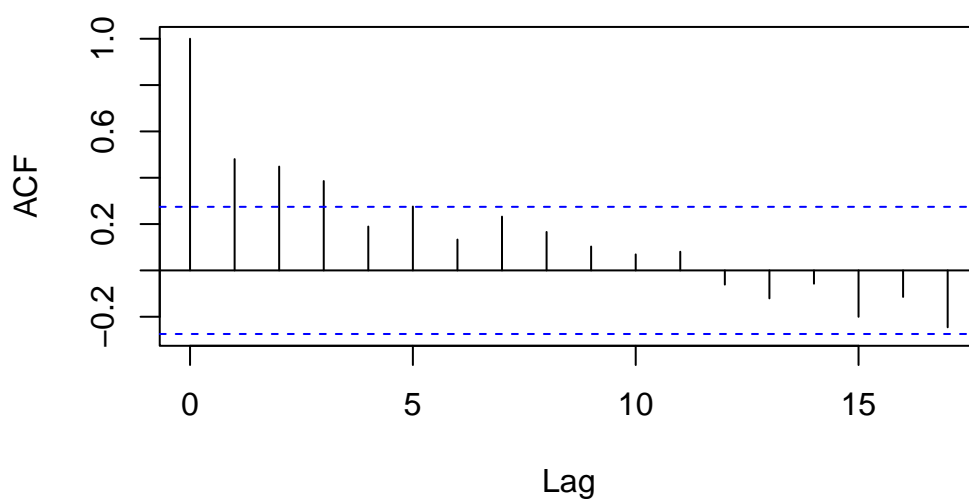
```
# 2nd Time Series  
# ACF of ts2  
acf(ts2$x)
```

### Series ts2\$x



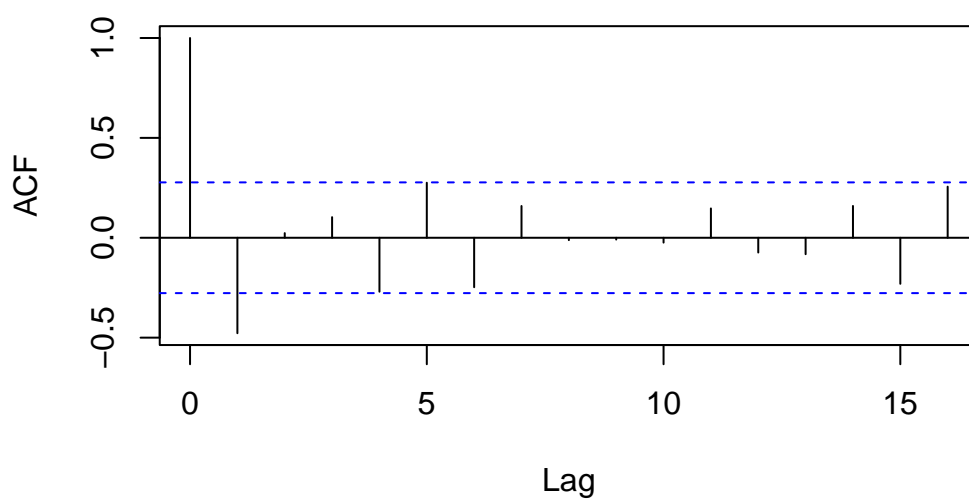
```
# ACF of Detrended by Regression  
acf(detrended_ts2)
```

### Series detrended\_ts2



```
# ACF of Detrended by Differencing  
acf(diff_ts2)
```

### Series diff\_ts2





---

## Part D

---

For the two different sets of data,  $ts1$  and  $ts2$ , we found that both sets had a lot of autocorrelation up to the 10th previous lag, which means they weren't really stable over time. Trying to smooth out the trends in  $ts1$  made it even less stable, but doing the same for  $ts2$  actually made it fit better within expected patterns, showing that removing trends can really help make the data more predictable. The most effective technique we used was differencing, which made both datasets much more stable and easier to analyze. This was especially true for the first couple of points in the series, which showed a strong connection after we applied differencing. This tells us that choosing the right method to stabilize our data is super important, and differencing seems to be a great choice, although we need to be careful not to overdo it and make our data look weird.