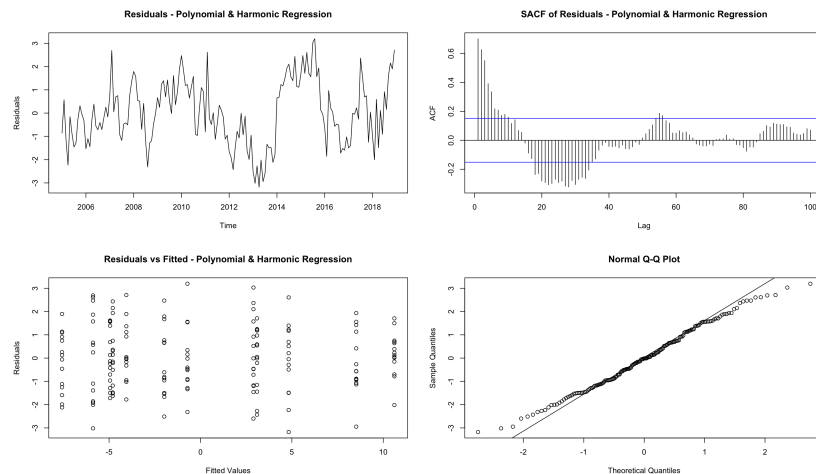
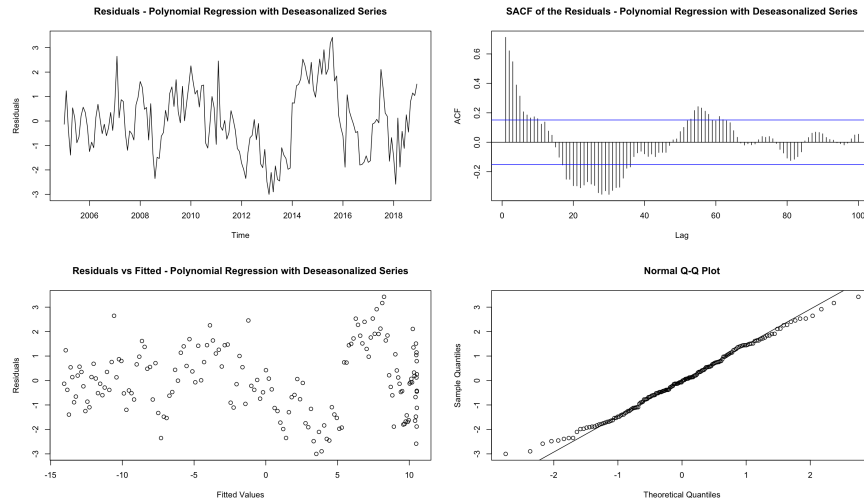


Time plot 을 통해, 주어진 월 단위 시계열 데이터에는 주기 12 의 계절성과 증가하는 추세가 존재함을 확인할 수 있다. Time plot 만으로는 형태를 확신할 수 없지만, 주어진 데이터의 추세는 직선 또는 2, 3 차 곡선의 형태를 가지고 있다. 더불어, 데이터의 이분산성은 존재하지 않으며, 이로 인하여 분산안정화 변환은 불필요하다. SACF 에서는 lag = 50 정도까지는 양의 상관관계가, 그 이후에는 대부분 음의 상관관계가 관찰되며, 특히 초기의 lag 에서는 아주 강한 양의 상관관계가 나타난다. SACF 상에서 12 개의 lag 마다 증가-감소의 패턴이 관찰되는 것을 보아 주어진 데이터에 주기 12 의 계절성이 존재하는 것을 다시 한번 확인할 수 있고, SACF 가 패턴을 가지면서도 선형적으로 감소하는 것으로 보아 주어진 데이터가 비정상 시계열임을 확인할 수 있다. 종합하여 보면, 주어진 데이터는 증가하는 추세와 더불어 주기 12 의 계절성이 존재하는 비정상 시계열이고, 따라서 이를 정상 시계열로 바꾸어주기 위해 추세를 추정하여 데이터에서 제거해야 한다.

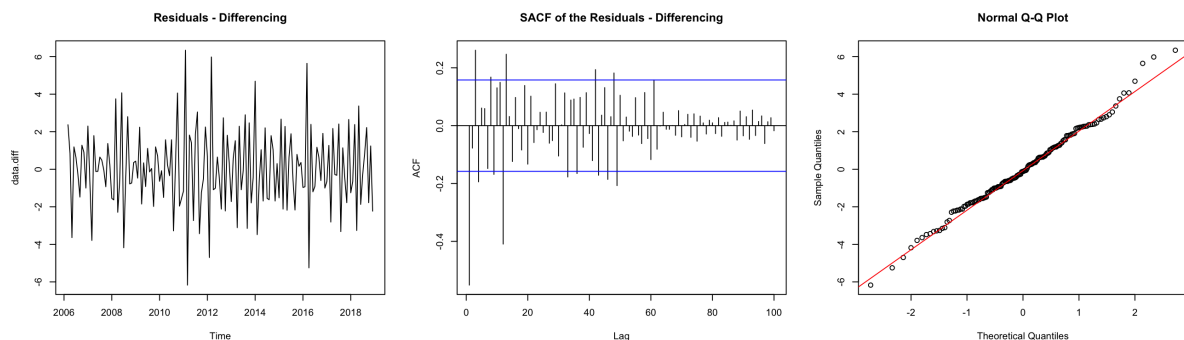
추세와 계절성을 제거하기 위해 회귀, 평활, 차분의 방법을 모두 사용해본 후 residual plot 및 residual 의 SACF 를 확인하여 가장 적합한 방법을 선택하였다.



우선, polynomial regression 으로 추세를 추정하고, harmonic regression 으로 계절성을 추정한 뒤 추정한 추세와 계절성을 데이터에서 제거한 후의 결과는 위와 같다. Polynomial regression 에서는 다항식의 차수를 1 차부터 4 차까지 조정하여 회귀식을 적합하였고, residual plot 상에서 추세가 가장 잘 제거되며 AIC 값도 1079.755 로 가장 낮았던 3 차를 최종 차수로 선택하였다. 4 차 다항식을 적합하는 경우에는 3 차에 비해 유의미한 AIC 값 변화가 없었고, 더불어 적합한 곡선의 형태 역시 3 차와 차이가 없어 3 차 다항식을 최종 선택하였다. 그 후에는 stepwise selection 방법을 이용하여 sin-cos 항의 개수를 조정된 harmonic regression 를 적합하여 계절성을 제거하였다. 이 때 사용된 항들은 costerm1 + sinterm1 + costerm2 + sinterm2 + costerm3 + sinterm3 + costerm4 + sinterm4 + sinterm5 + costerm6 이며, 이 때의 AIC 값은 608.4239 로 가장 작았다. 보통 sin-cos 쌍의 개수를 1 개부터 4 개까지로 설정하나, stepwise selection 을 사용하여 위와 같은 식을 적합했을 때의 AIC 값이 유의미하게 감소하여 stepwise selection 을 이용하여 harmonic regression 을 시행하였다. 그러나, 위의 결과를 보면 알 수 있듯이, polynomial regression 과 harmonic regression 으로 추정한 추세와 계절성을 제거한 residual 의 plot 상에서 여전히 추세가 관찰된다. 더불어, 해당 residual 의 SACF 에서도 여전히 선형적으로 감소하는 형태가 나타나는 것으로 보아 데이터의 추세가 완전히 제거되지 않아 residual 이 여전히 비정상 시계열임을 확인할 수 있다. 따라서, regression 방법은 주어진 데이터를 정상 시계열로 변환하기 위한 적합한 방법이 아니다.



다음으로는 smoothing-based classical decomposition algorithm 을 이용하여 추세와 계절성을 추정하였고, 이를 주어진 데이터에서 제거한 결과는 위와 같다. 우선, MA filter 를 이용하여 추세를 먼저 추정하였다. LOOCV 방법으로 구한 bandwidth = 39 를 이용하여 추세를 추정하였고, 그 결과 polynomial regression 과 비슷하지만 조금 더 복잡한 형태의 추세를 추정할 수 있었다. 추세가 제거된 데이터를 이용하여 seasonal smoothing 을 시행하였고, 이 때 추정된 계절성을 다시 원래의 데이터에서 빼준 후 polynomial regression 을 이용하여 최종적으로 추세를 추정한 후 제거해주었다. Polynomial regression 의 경우, 앞의 회귀 방법에서 다항식의 차수를 3 으로 두는 것이 가장 적합했고, MA filter 를 이용하여 추정한 곡선 역시 3 차 곡선의 형태와 비슷하였기 때문에 다항식의 차수를 3 으로 설정하였다. 그러나, residual plot 상에서 여전히 추세가 나타나고, residual 의 SACF 에서도 선형적으로 감소하는 형태가 여전히 존재하였다. 따라서, 앞의 회귀 방법의 결과와 동일하게 데이터의 추세가 완전히 제거되지 않아 residual 이 여전히 비정상시계열이며, 이로 인하여 smoothing 방법 역시 주어진 데이터를 정상시계열로 변환하기 위한 적절한 방법이 아니라는 결론을 내릴 수 있다.



비정상시계열을 정상시계열로 변환하기 위한 마지막 방법으로는 차분을 시도하였고, 그 결과는 위와 같다. 우선, lag 를 데이터의 주기인 12 로 두어 seasonal differencing 을 시행하였다. 그 후, 데이터에 존재하는 3 차 추세를 제거하기 위해 차분을 시행하였는데, 이 때 이미 seasonal differencing 에서 1 차 차분이 진행되었기 때문에 seasonal differencing 이후 차분을 두 번 더 진행해 주었다. 그 결과, residual plot 상에서 회귀 및 평활 방법에서 나타났던 추세가 나타나지 않았으며, 데이터의 추세 역시 완전히 제거되었다. Residual 의 SACF 에서도 회귀와 평활 방법에서 관찰되었던 선형적으로 감소하는 패턴이 나타나지 않았다. 마지막으로, normal-QQ plot 상에서도 residual 이 Gaussianity 를 만족함이 확인되었다. 이를 통해 주어진 비정상시계열 데이터를 정상시계열로 변환하기 위한 가장 적합한 방법은 차분이라고 결론을 내렸다.

차분 후의 residual 은 기존의 데이터에 존재하던 추세와 계절성이 완전히 제거되었고, 이분산성이 존재하지 않으며 SACF 상에서도 패턴이 관찰되지 않는다. 따라서, 차분 후의 residual 은 약정상성을 만족하는 정상시계열이다. 그런데, 위의 residual 에 대한 SACF 를 확인하면 lag = 1 에서 강한 음의 상관관계가 관찰되고, lag=20 이내에서도 양의 상관관계 및 음의 상관관계가 관찰된다. 따라서, residual 간 dependency 가 존재할 수 있다고 판단하였고, 만약 dependency 가 존재한다면 추후 이 정보를 모델링 해 주어야 하기 때문에 차분을 통해 구한 residual 이 IID noise 인지를 판단하기 위한 검정을 시행하였다.

Null hypothesis: Residuals are iid noise.			
Test	Distribution	Statistic	p-value
Ljung-Box Q	$Q \sim \text{chisq}(20)$	137.25	0 *
McLeod-Li Q	$Q \sim \text{chisq}(20)$	41.72	0.003 *
Turning points T	$(T-101.3)/5.2 \sim N(0,1)$	108	0.2
Diff signs S	$(S-76.5)/3.6 \sim N(0,1)$	77	0.8894
Rank P	$(P-5890.5)/320 \sim N(0,1)$	5796	0.7678

Residual 이 IID noise 인지에 대한 검정 결과는 위와 같다. 우선, Diff signs 및 Rank test 의 경우에는 선형 추세를 잡는데 더 강하므로 주어진 데이터와는 맞지 않는다고 판단하여 Ljung-Box, McLeod-Li 그리고 Turning points test 의 결과를 신뢰해야 한다고 판단하였다. 이 3 개의 검정 중 2 개의 검정인 Ljung-Box test, McLeod-Li test 에서 residual 이 IID noise 가 아니라는 결론이 나왔다. 따라서, 차분을 통해 구한 residual 은 IID noise 가 아니며, 정상시계열임에도 불구하고 여전히 dependency 가 존재하기 때문에 이후 이 dependency 구조를 모델링 해 주어야 한다. 이후 데이터가 Gaussianity 를 만족하는지의 여부에 대한 검정을 시행하였다. 이를 위해 Shapiro-wilk, Anderson-Darling, Cramer-von Mises, Kolmogorov-Smirnov 그리고 Jarque Bera test 의 총 5 가지의 검정을 시행하였으며, 각각의 p-value 는 0.5389, 0.5309, 0.6664, 0.689, 0.337 로 귀무가설이 기각되지 않았다. 따라서, 차분을 통해 구한 residual 은 Gaussianity 를 만족한다는 결론을 내렸다.

정리하자면, 주어진 데이터는 3 차 곡선 형태로 증가하는 추세가 존재하고, 주기 12 의 계절성이 존재하는 비정상시계열이다. 이를 정상시계열로 변환하기 위한 가장 적합한 방법은 차분이며, 우선 주기 12 의 계절 차분을 먼저 시행한 후 추세를 제거해 주기 위해 차분을 두 번 더 시행해 주어야 한다. 차분 후의 residual 은 추세와 계절성이 존재하지 않고, 분산이 일정한 정상시계열이다. 그러나, 차분 후의 residual 은 정상성을 만족하지만 residual 간의 dependency 가 존재하며, 따라서 IID noise 가 아니므로 이후 추가적인 모델링이 필요하다. 더불어, 정규성 검정을 통해 차분 후의 residual 이 Gaussianity 를 만족함이 확인되었다.

## [R code]

```
#####  
# Basic Settings #  
#####  
  
setwd("/Users/hyungyeonhong/Desktop/TSA_exam1")  
rm(list = ls(all = TRUE))  
source("TS-library.R")  
  
# Jan 2005 to Dec, 2018, hence it contains total 168 monthly data.  
data = read.csv("2022practice1.csv", header = F)  
data = data$V2  
  
data = ts(data, start = c(2005, 1), end = c(2018, 12), frequency = 12)  
n = length(data)  
  
#####  
# (a) Time plot, correlograms (ACF) and discuss key features of the data #  
#####  
  
par(mfrow = c(1, 2))  
  
plot.ts(data)  
title("Time Plot of the Data")  
  
acf2(data, lag = 100)  
title("SACF of the Data")  
  
#####  
# (b) Removing trend and seasonality #  
#####  
  
#####  
# Regression #  
#####  
  
#####  
# Removing Trend: Polynomial Regression #  
#####  
  
x = seq(from = 1, to = n, by = 1)  
x2 = x^2  
x3 = x^3  
  
out.polynomial1 = lm(data ~ 1 + x)  
out.polynomial2 = lm(data ~ 1 + x + x2)  
out.polynomial3 = lm(data ~ 1 + x + x2 + x3)  
  
AIC(out.polynomial1) # 1083.086  
AIC(out.polynomial2) # 1081.771  
AIC(out.polynomial3) # 1079.755  
  
par(mfrow = c(2, 2))  
  
plot.ts(data)  
lines(x = as.vector(time(data)), y = out.polynomial3$fitted.values, col = "red")  
title("Estimated Trend - Polynomial Regression")  
  
plot(x = as.vector(time(data)), y = out.polynomial3$residuals, type = "l")  
title("Residuals - Polynomial Regression")  
  
acf2(out.polynomial3$residuals)  
title("SACF of the Residuals - Polynomial Regression")  
  
qqnorm(out.polynomial3$residuals)  
qqline(out.polynomial3$residuals)  
  
out.polynomial = lm(data ~ 1 + x + x2 + x3)  
  
#####  
# Removing Seasonality: Harmonic Regression #  
#####  
  
t = 1:n  
d = 12  
  
f1 = n/d  
f2 = 2*f1  
f3 = 3*f1  
f4 = 4*f1  
f5 = 5*f1  
f6 = 6*f1  
  
costerm1 = cos(f1*2*pi/n*t); sinterm1 = sin(f1*2*pi/n*t);  
costerm2 = cos(f2*2*pi/n*t); sinterm2 = sin(f2*2*pi/n*t);  
costerm3 = cos(f3*2*pi/n*t); sinterm3 = sin(f3*2*pi/n*t);  
costerm4 = cos(f4*2*pi/n*t); sinterm4 = sin(f4*2*pi/n*t);  
costerm5 = cos(f5*2*pi/n*t); sinterm5 = sin(f5*2*pi/n*t);  
costerm6 = cos(f6*2*pi/n*t); sinterm6 = sin(f6*2*pi/n*t);
```

```

# Setting k = 1
out.harmonic1 = lm(out.polynomial$residuals ~ 1 + costerm1 + sinterm1)
summary(out.harmonic1)

# Setting k = 2
out.harmonic2 = lm(out.polynomial$residuals ~ 1 + costerm1 + sinterm1 + costerm2 + sinterm2)
summary(out.harmonic2)

# Setting k = 3
out.harmonic3 = lm(out.polynomial$residuals ~ 1 + costerm1 + sinterm1 + costerm2 + sinterm2 + costerm3 + sinterm3)
summary(out.harmonic3)

# Setting k = 4
out.harmonic4 = lm(out.polynomial$residuals ~ 1 + costerm1 + sinterm1 + costerm2 + sinterm2 + costerm3 + sinterm3 + costerm4 + sinterm4)
summary(out.harmonic4)

AIC(out.harmonic1) # 815.9302
AIC(out.harmonic2) # 742.3954
AIC(out.harmonic3) # 728.71
AIC(out.harmonic4) # 714.2348

# Variable selection to choose k: stepwise selection
out.total = lm(out.polynomial$residuals ~ 1 + costerm1 + sinterm1 + costerm2 + sinterm2 + costerm3 + sinterm3 +
               costerm4 + sinterm4 + costerm5 + sinterm5 + costerm6 + sinterm6)

step(out.total, direction = "both")

out.stepwise = lm(out.polynomial$residuals ~ costerm1 + sinterm1 +
                  costerm2 + sinterm2 + costerm3 + sinterm3 + costerm4 + sinterm4 +
                  sinterm5 + costerm6)

AIC(out.stepwise) # 608.4239

par(mfrow = c(3, 2))

plot(x = as.vector(time(data)), y = out.polynomial$residuals, type = "l")
lines(x = as.vector(time(data)), y = out.harmonic1$fitted.values, col = "salmon")
title("Harmonic Regression with k = 1")

plot(x = as.vector(time(data)), y = out.polynomial$residuals, type = "l")
lines(x = as.vector(time(data)), y = out.harmonic2$fitted.values, col = "lightblue")
title("Harmonic Regression with k = 2")

plot(x = as.vector(time(data)), y = out.polynomial$residuals, type = "l")
lines(x = as.vector(time(data)), y = out.harmonic3$fitted.values, col = "lightgreen")
title("Harmonic Regression with k = 3")

plot(x = as.vector(time(data)), y = out.polynomial$residuals, type = "l")
lines(x = as.vector(time(data)), y = out.harmonic4$fitted.values, col = "orange")
title("Harmonic Regression with k = 4")

plot(x = as.vector(time(data)), y = out.polynomial$residuals, type = "l")
lines(x = as.vector(time(data)), y = out.stepwise$fitted.values, col = "pink")
title("Harmonic Regression with Stepwise Selection")

par(mfrow = c(2, 2))

plot(x = as.vector(time(data)), y = out.stepwise$residuals, type = "l", xlab = "Time", ylab = "Residuals")
title("Residuals - Polynomial & Harmonic Regression")

acf2(out.stepwise$residuals, lag = 100)
title("SACF of Residuals - Polynomial & Harmonic Regression")

plot(out.stepwise$fitted.values, out.stepwise$residuals, xlab = "Fitted Values", ylab = "Residuals")
title("Residuals vs Fitted - Polynomial & Harmonic Regression")

qqnorm(out.stepwise$residuals)
qqline(out.stepwise$residuals)

#####
# Smoothing : Smoothing-Based Classical Decomposition Algorithm #
#####

#####
# Removing Trend: MA filter #
#####

h.ma = optimize(f = ma.cv, interval = c(5, n/2), Y = data, l = 1, tol = .Machine$double.eps^0.25)
h.optimal = round(h.ma$minimum)
out.ma = smooth.ma(data, h.optimal)

par(mfrow = c(2, 2))

plot.ts(data)
lines(x = as.vector(time(data)), out.ma, col = "red")
title("Estimated Trend - MA Filter")

plot.ts(data - out.ma)
title("Residuals - MA Filter")

acf2(data - out.ma)

```

```

title("SACF of the Residuals Obtained by MA Filter")

qqnorm(data ~ out.ma)
qqline(data ~ out.ma, col="red")

#####
# Removing Seasonality: Seasonal Smoothing #
#####

temp.detrrend = data ~ out.ma

season.avg = season(temp.detrrend, d = 12)

par(mfrow = c(2, 2))

plot.ts(temp.detrrend)
lines(x = as.vector(time(temp.detrrend)), y = season.avg + mean(temp.detrrend), col = "red")
title("Seasonal Smoothing")

plot(x = as.vector(time(temp.detrrend)), y = temp.detrrend ~ season.avg ~ mean(temp.detrrend), type = "T")
title("Residuals ~ MA filter & Seasonal Smoothing")

acf2(temp.detrrend ~ season.avg ~ mean(temp.detrrend), lag = 100)
title("SACF of the Residuals ~ MA filter & Seasonal Smoothing")

qqnorm(temp.detrrend ~ season.avg ~ mean(temp.detrrend))
qqline(temp.detrrend ~ season.avg ~ mean(temp.detrrend))

#####
# Re-estimating Trend: Polynomial Regression #
#####

data.deseasonalized = data ~ season.avg ~ mean(data)
plot.ts(data.deseasonalized)

x = seq(from = 1, to = n, by = 1)
x2 = x^2
x3 = x^3

out.smooth.reg = lm(data.deseasonalized ~ 1 + x + x2 + x3)
summary(out.smooth.reg)

par(mfrow = c(1, 1))

plot.ts(data.deseasonalized)
lines(x = as.vector(time(data.deseasonalized)), y = out.smooth.reg$fitted.values, col = "red")
title("Estimated Trend ~ Polynomial Regression with Deseasonalized Series")

par(mfrow = c(2, 2))

plot(x = as.vector(time(data.deseasonalized)), y = out.smooth.reg$residuals, type = "l",
      xlab = "Time", ylab = "Residuals")
title("Residuals ~ Polynomial Regression with Deseasonalized Series")

acf2(out.smooth.reg$residuals, lag = 100)
title("SACF of the Residuals ~ Polynomial Regression with Deseasonalized Series")

plot(x = out.smooth.reg$fitted.values, y = out.smooth.reg$residuals,
      xlab = "Fitted Values", ylab = "Residuals")
title("Residuals vs Fitted ~ Polynomial Regression with Deseasonalized Series")

qqnorm(out.smooth.reg$residuals)
qqline(out.smooth.reg$residuals)

#####
# Differencing #
#####

#####
# Removing seasonality by lag-d differencing #
#####

d = 12
season.diff = diff(data, lag = d)

par(mfrow = c(1, 3))

plot(season.diff, type = "l")
title("Residuals ~ Seasonal Differencing") # includes first order differencing

acf2(season.diff, lag = 100)
title("SACF of the Residuals ~ Seasonal Differencing")

qqnorm(season.diff)
qqline(season.diff)

#####
# Removing trend by differencing #
#####

data.diff = diff(diff(season.diff)) # apply diff() multiple times if needed

```

```
par(mfrow = c(1, 3))

plot.ts(data,diff)
title("Residuals - Differencing")

acf2(data,diff, lag = 100) # seasonality not fully removed
title("SACF of the Residuals - Differencing")

qqnorm(data,diff)
qqline(data,diff, col="red")

#####
# Test of Randomness #
#####

#####
# IID, WN Sequence #
#####

test(data,diff)

#####
# Gaussianity #
#####

shapiro.test(data,diff)

library(nortest)

ad.test(data,diff)
cvm.test(data,diff)
lillie.test(data,diff)

library(tseries)
jarque.bera.test(data,diff)
```