STA 137 Final Project

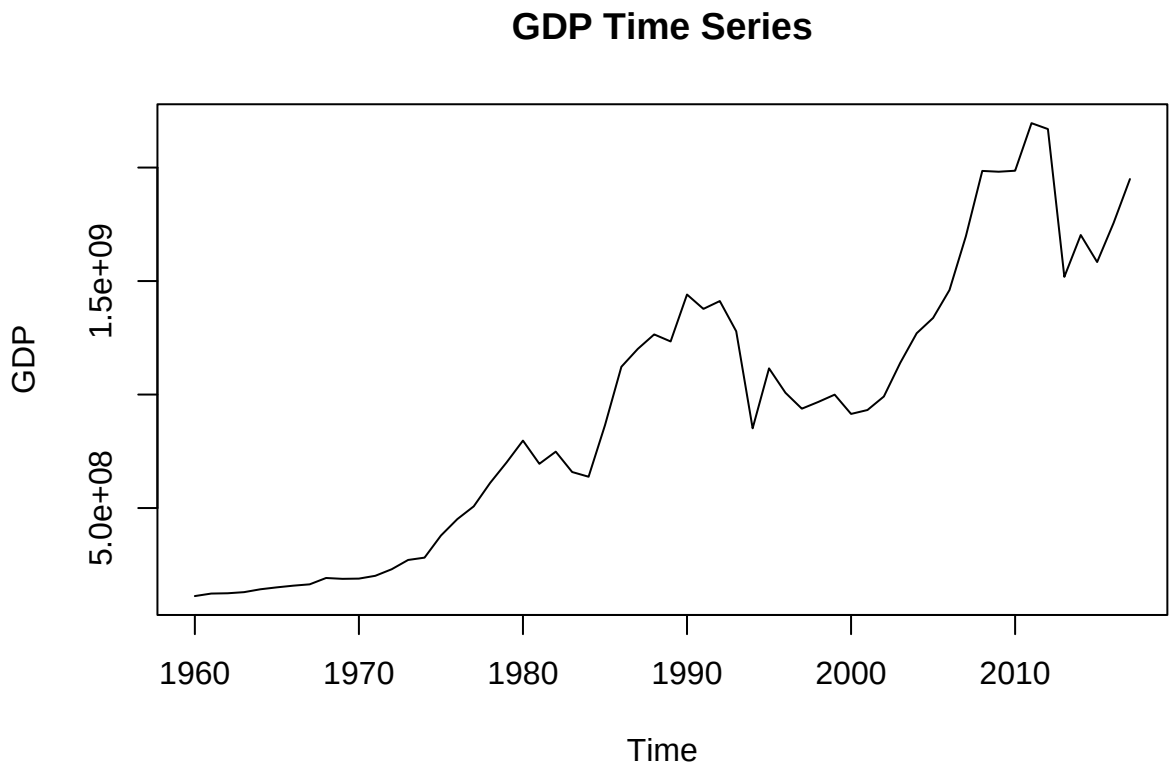
Quynh Trinh,

2025-06-04

GDP Time Series

Plot GDP Time Series & Observing

```
# Plot GDP
gdp_ts <- ts(finalPro_data$GDP, start = 1960, frequency = 1)

ts.plot(gdp_ts, main="GDP Time Series", ylab="GDP")
```



Summary:

- GDP time series has upward trend, this shows this is non-stationary
- It has peaks around every 10 year: 1980, 1990, 2010

Diagnostic GDP

Coefficients

```
#
model_gdp <- lm(GDP ~ Year + Growth + CPI + Imports + Exports + Population, data = finalPro_data)

summary(model_gdp)
```

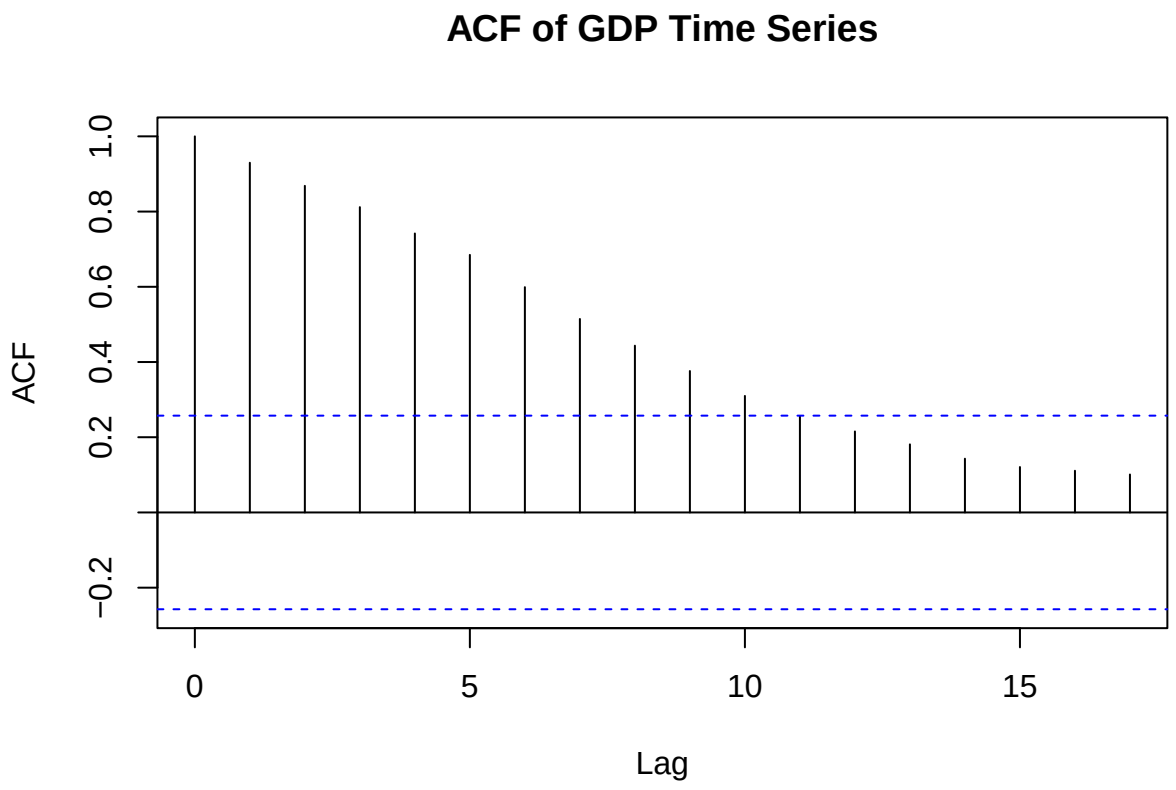
```
##
## Call:
## lm(formula = GDP ~ Year + Growth + CPI + Imports + Exports +
##     Population, data = finalPro_data)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -355647077 -123917222   18740212  114338509  328824034
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.704e+11  1.190e+11  -2.272   0.0307 *
## Year         1.394e+08  6.097e+07   2.286   0.0297 *
## Growth       6.543e+06  4.960e+06   1.319   0.1974
## CPI         -9.881e+06  4.121e+06  -2.398   0.0231 *
## Imports      1.590e+07  1.088e+07   1.461   0.1548
## Exports     -7.174e+07  1.240e+07  -5.784 2.89e-06 ***
## Population  -1.470e+03  7.621e+02  -1.928   0.0636 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 194300000 on 29 degrees of freedom
##   (22 observations deleted due to missingness)
## Multiple R-squared:  0.8316, Adjusted R-squared:  0.7967
## F-statistic: 23.87 on 6 and 29 DF,  p-value: 5.499e-10
```

(Note: default p-value = 0.1)

Significant: Year, CPI, Exports

Residuals

```
# Residuals diagnostics for GDP
acf(gdp_ts, main = "ACF of GDP Time Series")
```




```
#Ljung test
Box.test(gdp_ts)
```

```
##
##  Box-Pierce test
##
## data:  gdp_ts
## X-squared = 50.145, df = 1, p-value = 1.428e-12
```

ACF:

- ACF values are decreasing gradually and stay above significance bounds, this means the time series is non-stationary, it likely has trend

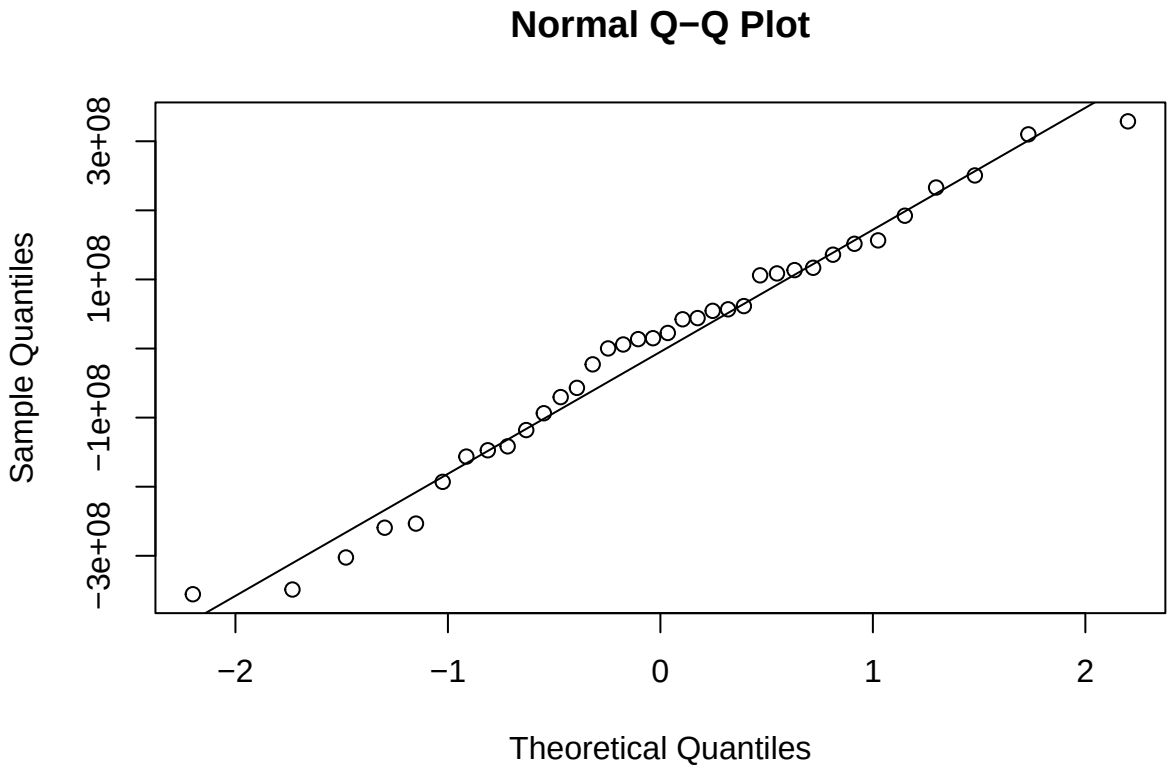
Ljung-test

- p-value = 1.428e-12 (very small), this means the residuals are not independent.

Normality and Constant Variance

```
resid_gdp <- residuals(model_gdp)
```

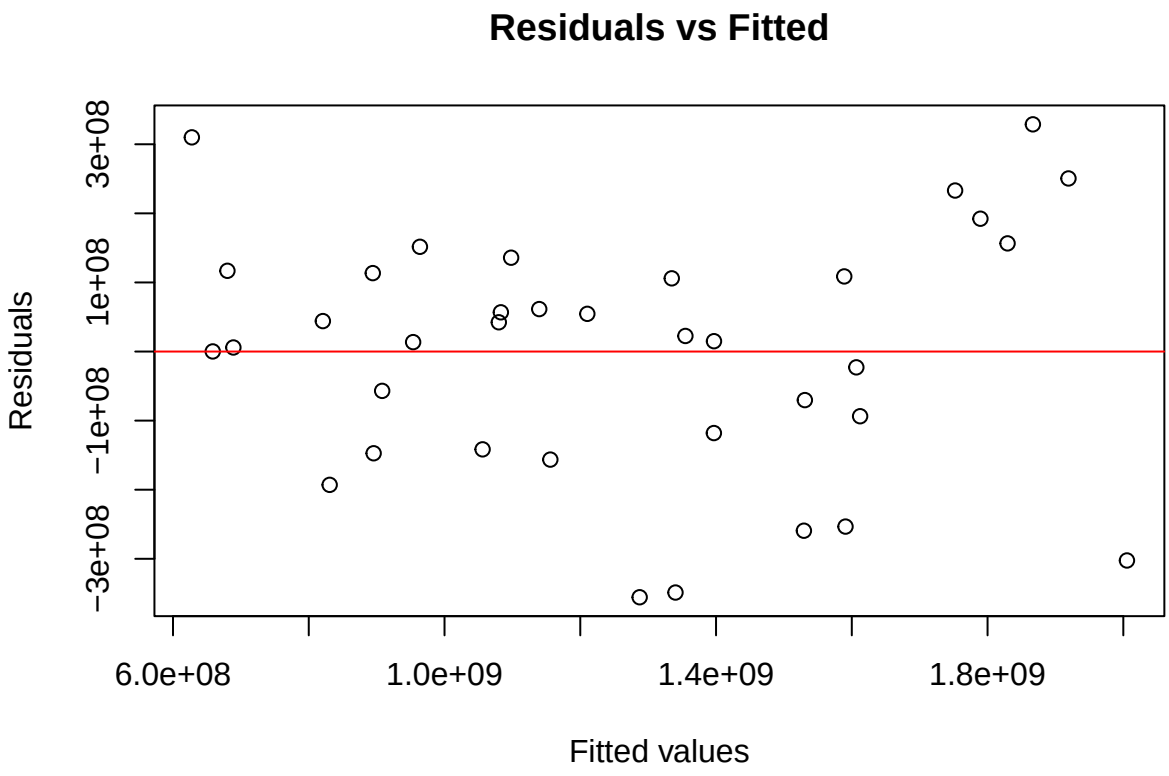
```r
# Check normality
qqnorm(model_gdp$residuals)
qqline(model_gdp$residuals)
```




```r
shapiro.test(resid_gdp) # Shapiro-Wilk test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid_gdp
## W = 0.97508, p-value = 0.5794
```

```r
# Check constant variance
plot(fitted(model_gdp), resid_gdp,
     xlab = "Fitted values", ylab = "Residuals",
     main = "Residuals vs Fitted")
abline(h = 0, col = "red") # Residuals vs Fitted plot
```




```
group <- ifelse(fitted(model_gdp) > median(fitted(model_gdp)), "High", "Low")
leveneTest(resid_gdp ~ group, center=median)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.

## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value  Pr(>F)
## group  1  6.5555 0.01507 *
##       34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Normality:

- The residuals are close to the line, this means the data seems to meet normality
- Shapiro: p-value = 0.5794. This means the data met normality

Constant Variance:

- Plot shows no patterns, this means the data seems to meet constant variance
- Brown-Forsythe test: p-value = 0.01507 < 0.05. This means the data not met constant variance. Thus, transformation is needed.

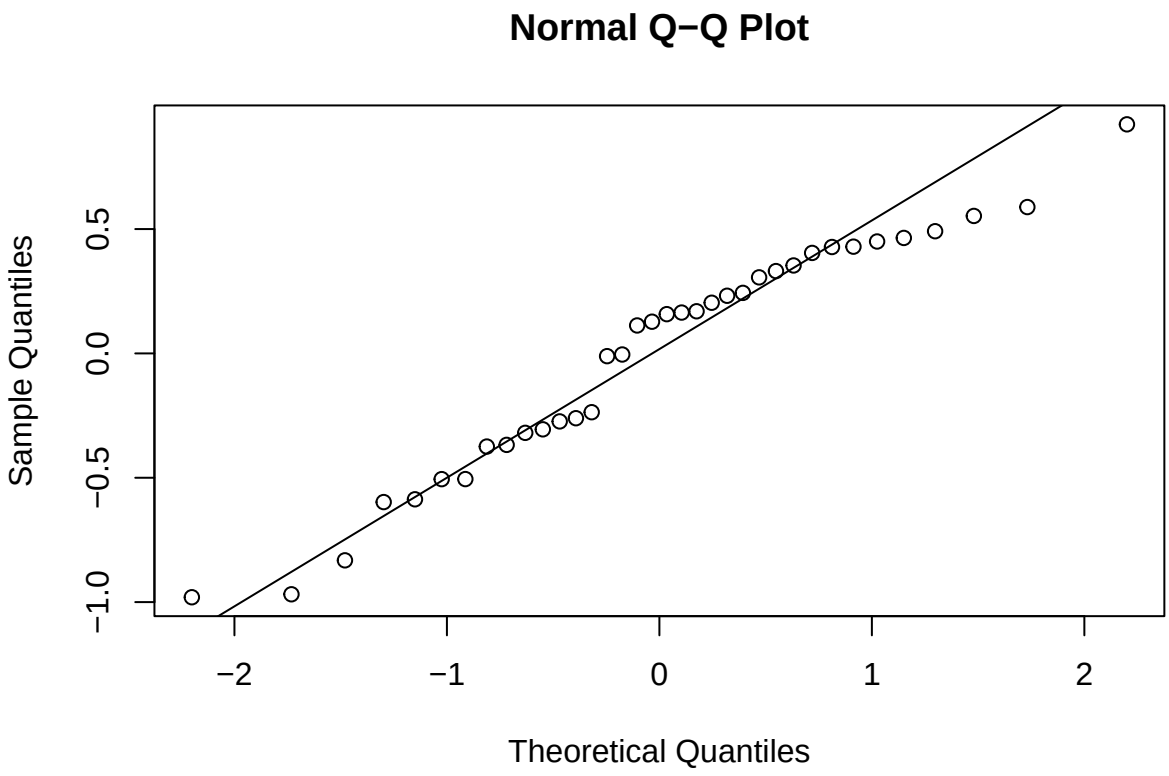
Box-Cox Transformation for GDP

Diagnostic for Tranformation

```
##
## Call:
## lm(formula = GDP_boxcox ~ Year + Growth + CPI + Imports + Exports +
##     Population, data = finalPro_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9805 -0.3315  0.1426  0.3662  0.9210
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.288e+02  3.168e+02  -2.616  0.01398 *
## Year         4.474e-01  1.623e-01   2.757  0.00999 **
## Growth       1.540e-02  1.320e-02   1.166  0.25295
## CPI         -2.445e-02  1.097e-02  -2.229  0.03373 *
## Imports      1.188e-02  2.898e-02   0.410  0.68487
## Exports     -1.901e-01  3.302e-02  -5.755 3.12e-06 ***
## Population  -5.099e-06  2.029e-06  -2.513  0.01777 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5172 on 29 degrees of freedom
##   (22 observations deleted due to missingness)
## Multiple R-squared:  0.847,  Adjusted R-squared:  0.8154
## F-statistic: 26.76 on 6 and 29 DF,  p-value: 1.41e-10
```

Significant coefficients: Year, CPI, Exports, Population

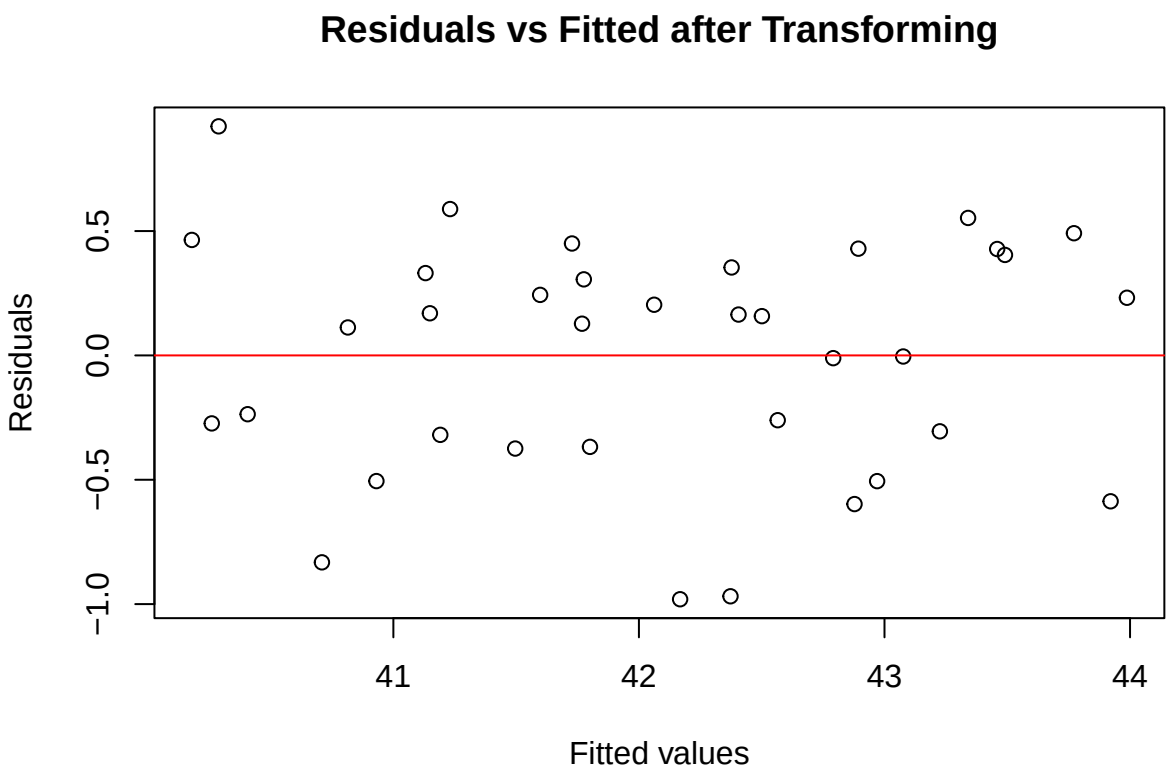
```
# Normality box-cox
resid_gdp_boxcox <- residuals(model_gdp_boxcox)

# Check normality
qqnorm(model_gdp_boxcox$residuals)
qqline(model_gdp_boxcox$residuals)
```




```
shapiro.test(resid_gdp_boxcox) # Shapiro-Wilk test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid_gdp_boxcox
## W = 0.95513, p-value = 0.1516
```

```
# Check constant variance
plot(fitted(model_gdp_boxcox), resid_gdp_boxcox,
     xlab = "Fitted values", ylab = "Residuals",
     main = "Residuals vs Fitted after Transforming")
abline(h = 0, col = "red") # Residuals vs Fitted plot
```




```
group_gdp_boxcox <- ifelse(fitted(model_gdp_boxcox) > median(fitted(model_gdp_boxcox)), "High", "Low")
leveneTest(resid_gdp_boxcox ~ group_gdp_boxcox, center=median)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  0.3973 0.5327
##       34
```

Normality and Constant Variance:

- Both assumptions are met after transforming

Find the best ARIMA model for GDP Using auto.arima()

```
arima_gdp <- auto.arima(gdp_ts)
arima_gdp
```

```
## Series: gdp_ts
## ARIMA(0,1,0) with drift
```

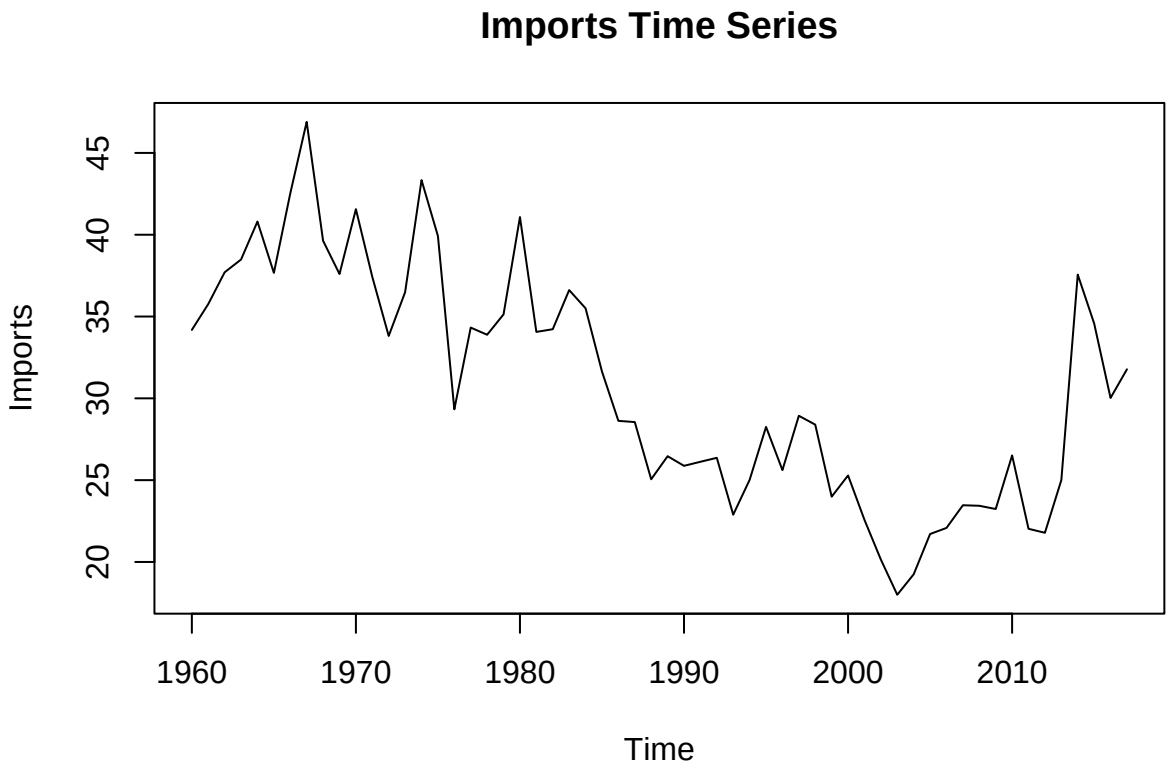
```
##
## Coefficients:
##          drift
##       32232562
## s.e.  19852873
##
## sigma^2 = 2.269e+16:  log likelihood = -1153.71
## AIC=2311.42   AICc=2311.64   BIC=2315.51
```

Suggested: ARIMA(0,1,0)

Imports Time Series

Plot Imports Time Series

```r
# Plot imports
imports_ts <- ts(finalPro_data$Imports, start = 1960, frequency = 1)

ts.plot(imports_ts, main="Imports Time Series", ylab="Imports")
```



Summary:

- The plot shows there is downward trend from 1960 to 2005, and increasing after that
- This means the Imports time series is non-stationary

Diagnostics Imports

Coefficients

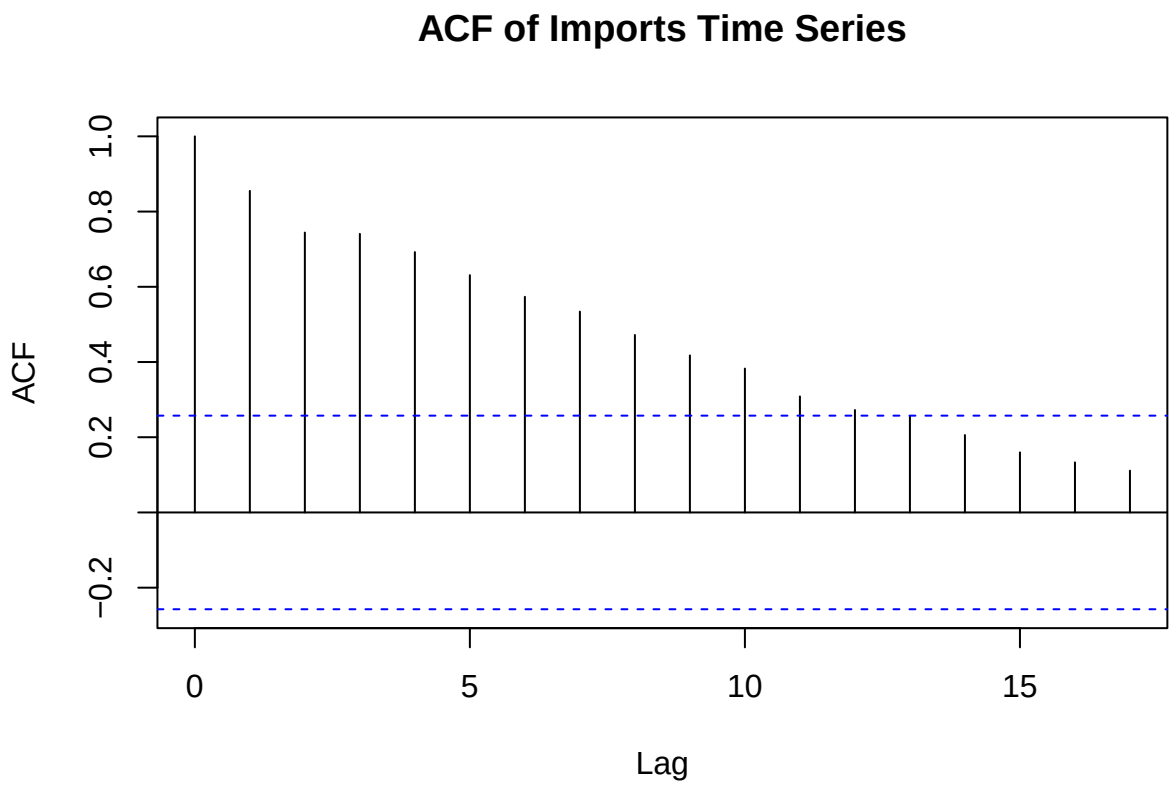
```
model_imports <- lm(Imports ~ Year + Growth + CPI + GDP + Exports + Population, data = finalPro_data)

summary(model_imports)
```

```
##
## Call:
## lm(formula = Imports ~ Year + Growth + CPI + GDP + Exports +
##     Population, data = finalPro_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8591 -1.7526 -0.4215  1.2413  8.9778
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.762e+02  2.121e+03   0.413  0.68251
## Year        -4.330e-01  1.088e+00  -0.398  0.69347
## Growth      -5.806e-02  8.339e-02  -0.696  0.49179
## CPI          2.359e-01  5.998e-02   3.933  0.00048 ***
## GDP          4.311e-09  2.951e-09   1.461  0.15484
## Exports      5.924e-01  2.788e-01   2.125  0.04226 *
## Population  -5.022e-06  1.330e-05  -0.378  0.70842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.199 on 29 degrees of freedom
##   (22 observations deleted due to missingness)
## Multiple R-squared:  0.7213, Adjusted R-squared:  0.6636
## F-statistic: 12.51 on 6 and 29 DF,  p-value: 6.306e-07
```

Significant: CPI, Exports

Residuals

```
# Residuals diagnostics for Imports
acf(imports_ts, main = "ACF of Imports Time Series")
```




```
#Ljung test
Box.test(imports_ts)
```

```
##
##  Box-Pierce test
##
## data:  imports_ts
## X-squared = 42.394, df = 1, p-value = 7.46e-11
```

ACF:

- ACF values decrease gradually and stay above significance bounds, this means the time series is non-stationary

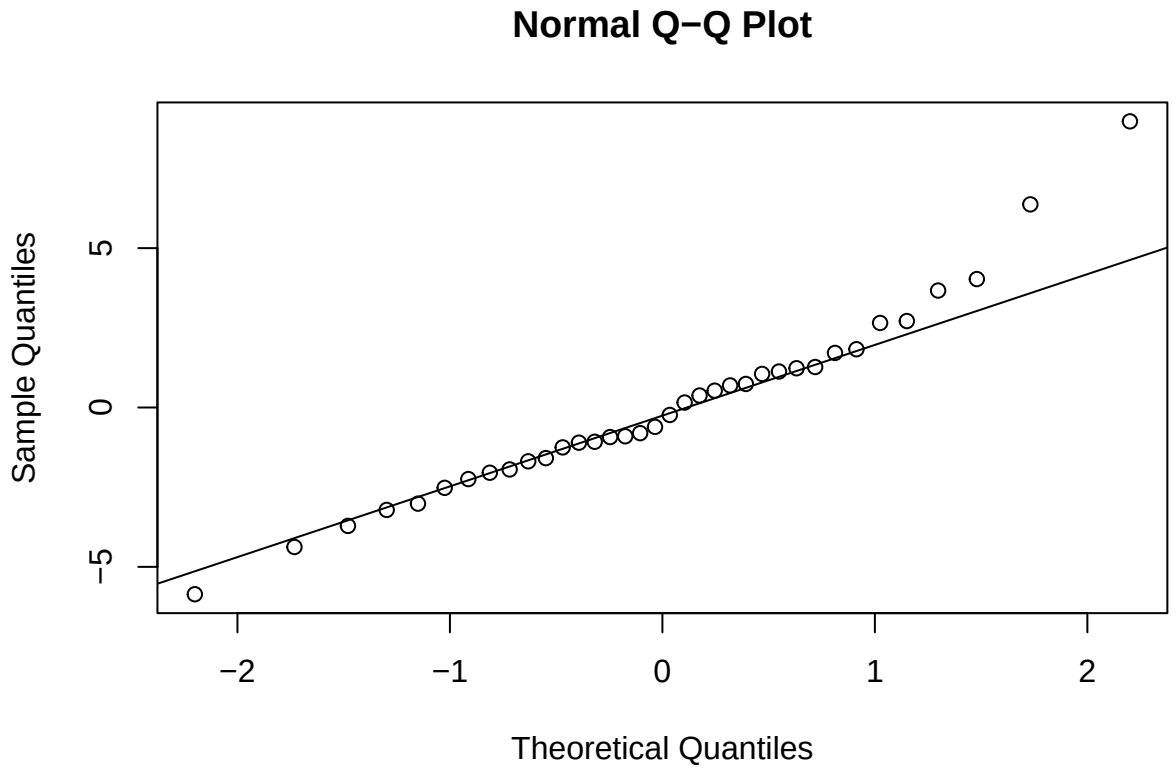
Ljung test:

- p-value = 7.46e-11, this means the residuals are not independent

Normality and Constant Variance

```
resid_imports <- residuals(model_imports)
```

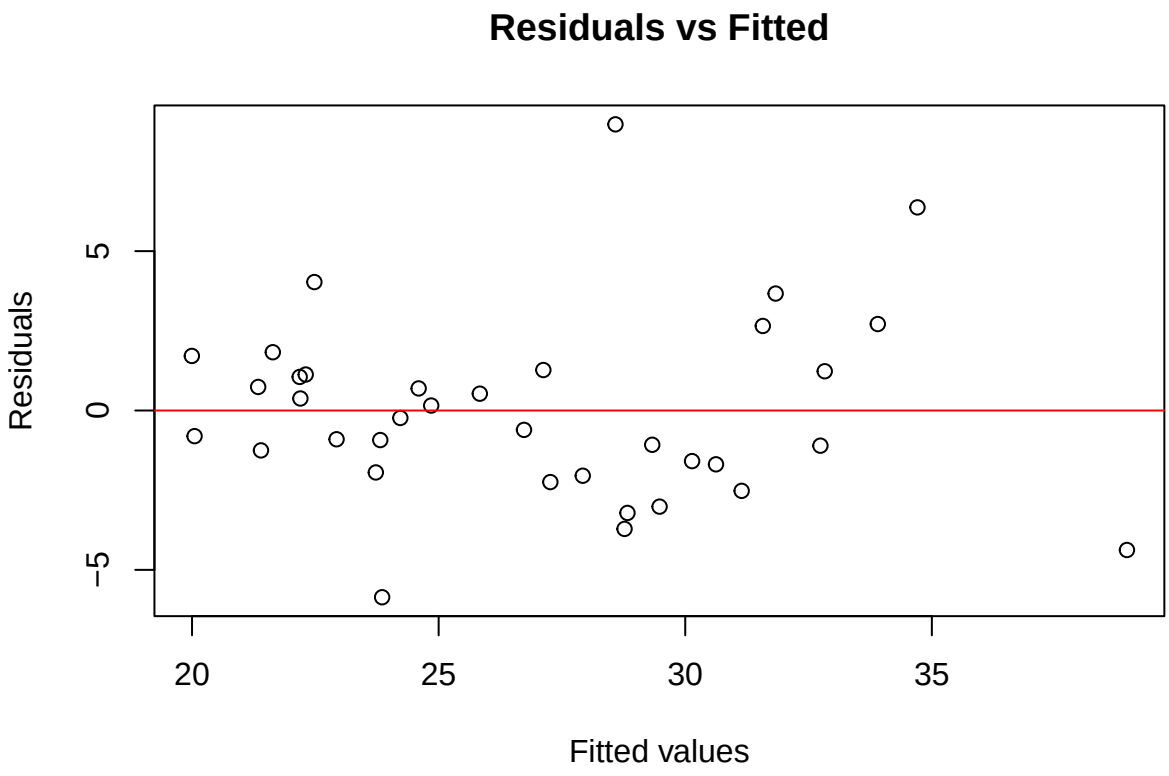
```r
# Check normality
qqnorm(model_imports$residuals)
qqline(model_imports$residuals)
```




```r
shapiro.test(resid_imports) # Shapiro-Wilk test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid_imports
## W = 0.95711, p-value = 0.1748
```

```r
# Check constant variance
plot(fitted(model_imports), resid_imports,
     xlab = "Fitted values", ylab = "Residuals",
     main = "Residuals vs Fitted")
abline(h = 0, col = "red") # Residuals vs Fitted plot
```




```
group_imports <- ifelse(fitted(model_imports) > median(fitted(model_imports)), "High", "Low")
leveneTest(resid_imports ~ group_imports, center=median)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value  Pr(>F)
## group  1   3.495 0.07019 .
##       34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Normality:

- Most residuals are close to the line except some points at the right end of the plot. The data might meet normality
- Shapiro test: p-value = 0.1748. This means the data met normality

Constant Variance:

- The plot shows some patterns; most points are at the left side and waving pattern. This means the data might not meet constant variance
- Brown test: 0.07019 > 0.05. This means the data met constant variance

Thus, there is no needed for transforming Imports model.

Find the Best ARIMA Model for Imports Using auto.arima()

```
arima_imports <- auto.arima(imports_ts)
arima_imports
```

```
## Series: imports_ts
## ARIMA(0,1,2)
##
## Coefficients:
##           ma1      ma2
##       -0.0463  -0.4473
## s.e.   0.1307   0.1361
##
## sigma^2 = 12.33:  log likelihood = -151.68
## AIC=309.37   AICc=309.82   BIC=315.5
```

Suggested: ARIMA(0,1,2)