GDP ARIMA

Andrew Jowe

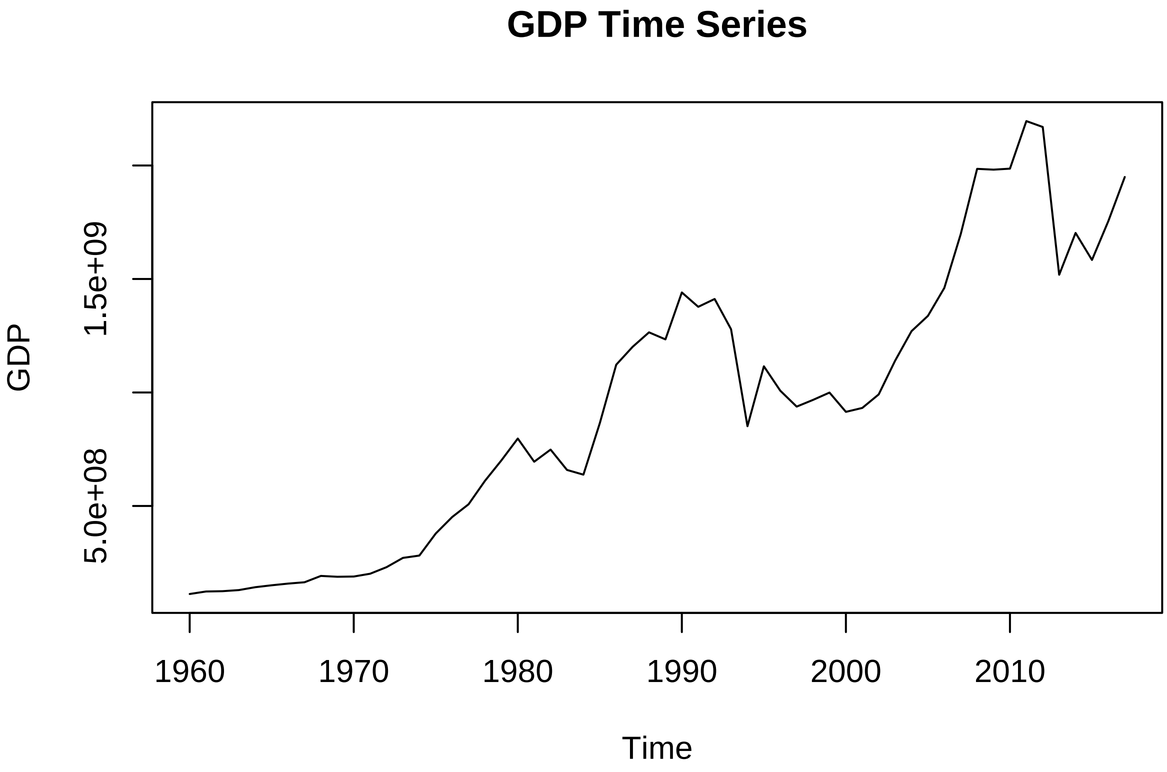
Col Removal

Keep Year, Imports, and GDP columns

```
finalPro_data <- finalPro_data[, c("Year", "GDP")]
```

Plot Time Series

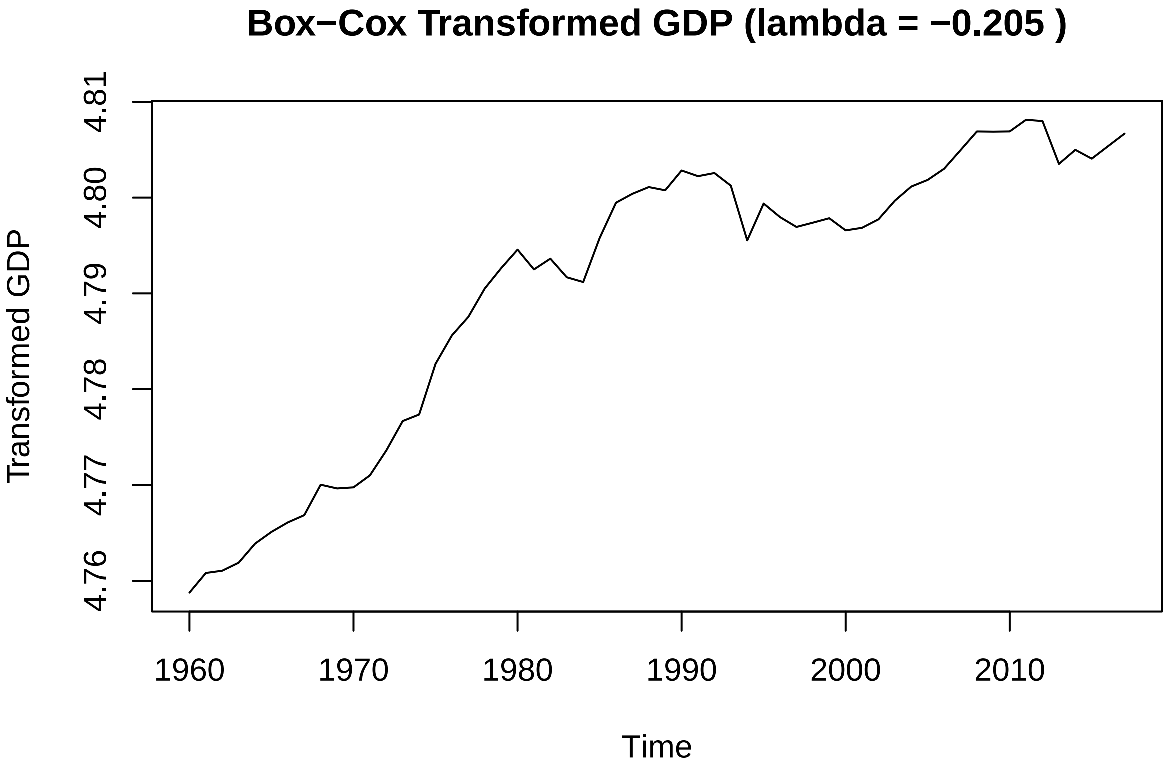
```
# Plot GDP
gdp_ts <- ts(finalPro_data$GDP, start = 1960, frequency = 1)

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

Transform

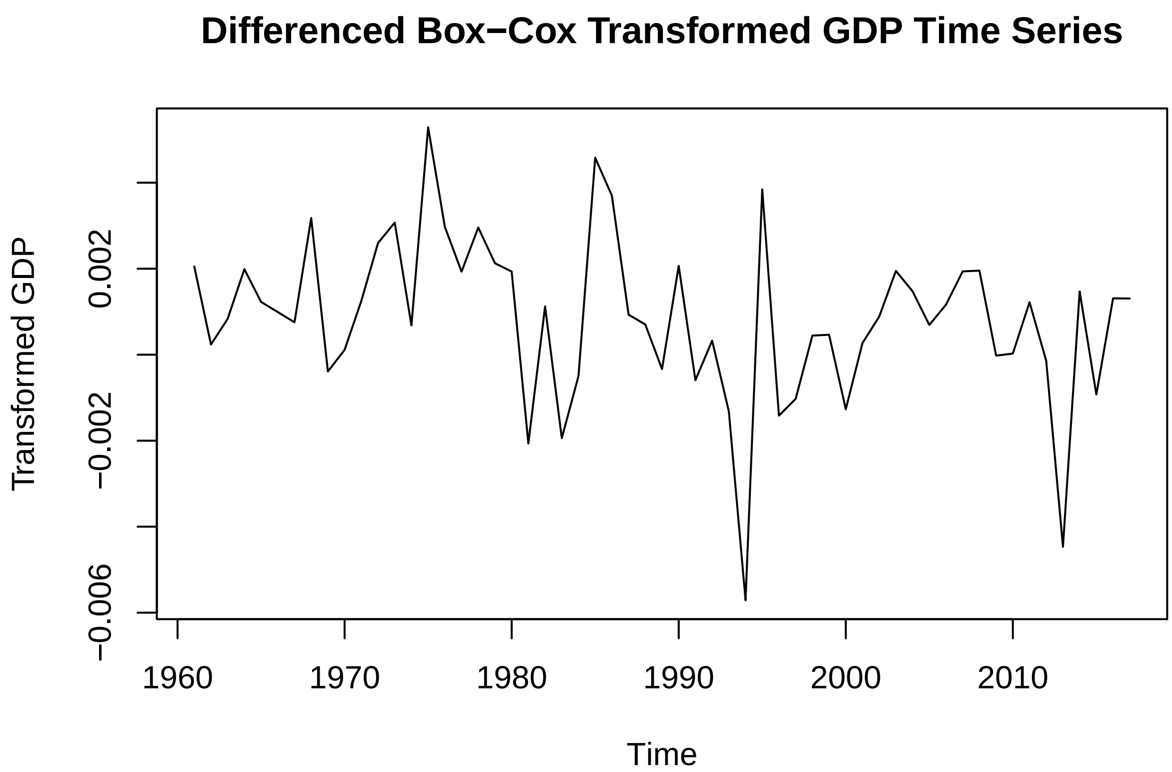
```r
# Box-Cox transform GDP
lambda <- BoxCox.lambda(gdp_ts)
boxcox_gdp_ts <- BoxCox(gdp_ts, lambda)
ts.plot(boxcox_gdp_ts, main = paste("Box-Cox Transformed GDP (lambda =", round(lambda, 3), ")"), ylab =
```



We tried log, but residuals not normal.

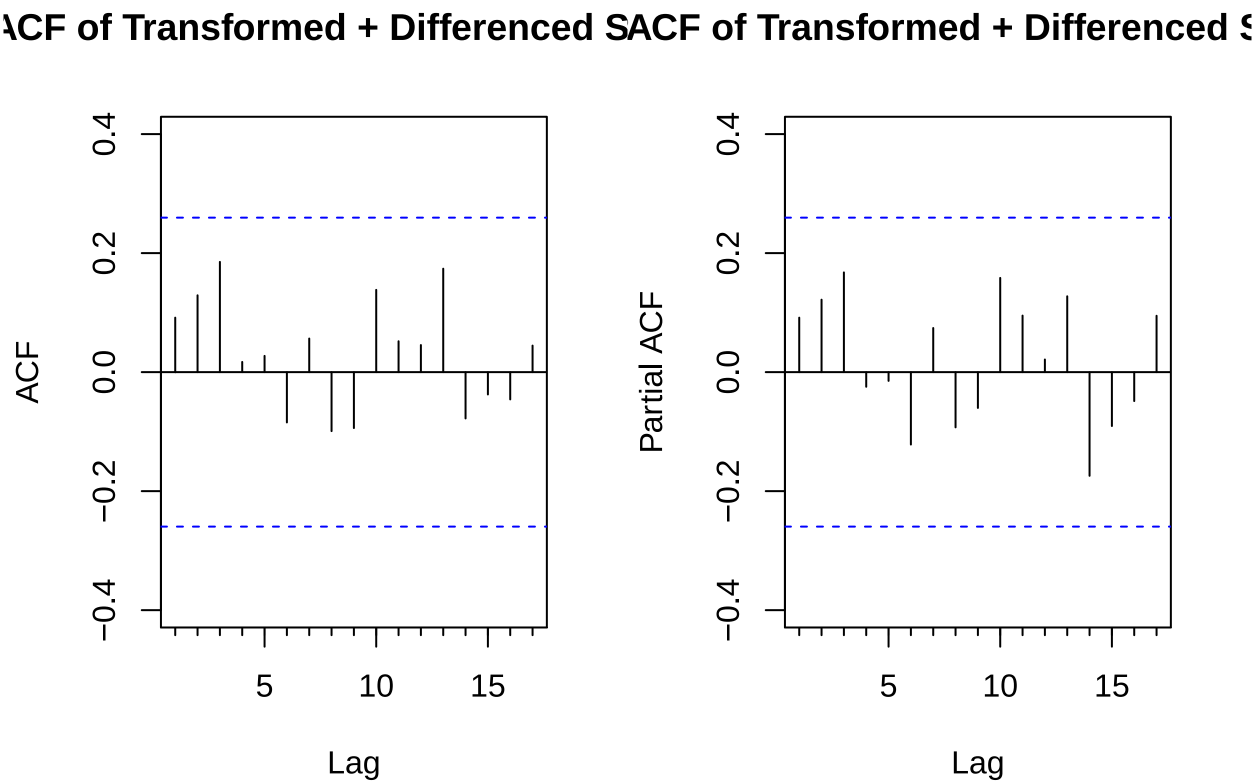
Differencing GDP

```r
diff_gdp_bc <- diff(boxcox_gdp_ts)

# Plot differenced Box-Cox GDP
ts.plot(diff_gdp_bc, main="Differenced Box-Cox Transformed GDP Time Series", ylab="Transformed GDP")
```



ACF / PACF plots

```
# ACF and PACF of the transformed and differenced series
par(mfrow = c(1, 2))
Acf(diff_gdp_bc, main = "ACF of Transformed + Differenced Series")
Pacf(diff_gdp_bc, main = "PACF of Transformed + Differenced Series")
```



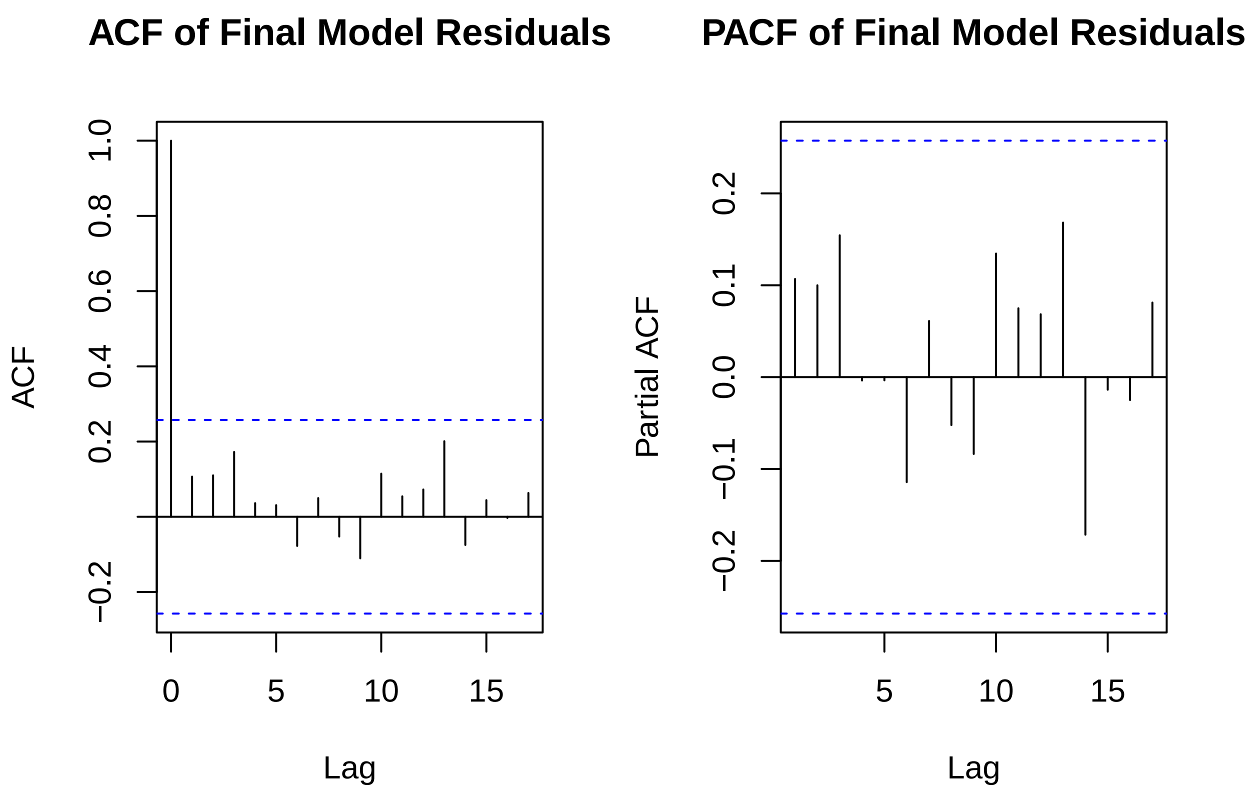

```r
par(mfrow = c(1, 1))
```

Modeling

```r
# Central African Republic GDP ARIMA Model
# Author: Om C

# Diagnostics on chosen model
final_model <- Arima(boxcox_gdp_ts, order = c(0,1,0), method = "ML")
print(final_model)
```

```
## Series: boxcox_gdp_ts
## ARIMA(0,1,0)
##
## sigma^2 = 4.727e-06:  log likelihood = 271.1
## AIC=-540.2   AICc=-540.12   BIC=-538.15
```

```r
residuals_final <- residuals(final_model)

# Residual ACF and PACF for final model
par(mfrow = c(1, 2))  # Side-by-side layout
acf(residuals_final, main = "ACF of Final Model Residuals")
pacf(residuals_final, main = "PACF of Final Model Residuals")
```




```
par(mfrow = c(1, 1))  # Reset layout

# Diagnostic Tests (Simplified)
cat("\nDiagnostic Tests (Simplified):\n")
```

```
##
## Diagnostic Tests (Simplified):
```

```
# 1. Portmanteau (Ljung-Box) Test for Autocorrelation
ljung <- Box.test(residuals_final, lag = 10, type = "Ljung-Box")
cat("Ljung-Box test p-value:", round(ljung$p.value, 4),
    ifelse(ljung$p.value > 0.05, "(PASS - residuals approx white noise)", "(FAIL)"), "\n")
```

```
## Ljung-Box test p-value: 0.8095 (PASS - residuals approx white noise)
```

```
# 2. Shapiro-Wilk Test for Normality
# this does not violate model assumptions, but it violates confidence interval assumptions
shapiro <- shapiro.test(residuals_final)
cat("Shapiro-Wilk test p-value:", round(shapiro$p.value, 4),
    ifelse(shapiro$p.value > 0.05, "(PASS - approx. normal residuals)", "(FAIL)"), "\n")
```

```
## Shapiro-Wilk test p-value: 0.0449 (FAIL)
```

```
# STEP 3: Model Comparison
models <- list(
  "ARIMA(0,1,0)" = c(0,1,0), # This implies a random walk, which is obviously an underfit for our model
  "ARIMA(1,1,0)" = c(1,1,0),
  "ARIMA(0,1,1)" = c(0,1,1),
  "ARIMA(1,1,1)" = c(1,1,1),
  "ARIMA(2,1,0)" = c(2,1,0),
  "ARIMA(2,1,1)" = c(2,1,1),
```

```r
  "ARIMA(2,1,2)" = c(2,1,2),
  "ARIMA(1,1,2)" = c(1,1,2),
  "ARIMA(0,1,2)" = c(0,1,2)
)
results <- data.frame(Model=character(), AIC=numeric(), BIC=numeric(),
                      Ljung_Box_p=numeric(), stringsAsFactors=FALSE)

for(i in 1:length(models)) {
  fit <- Arima(boxcox_gdp_ts, order = models[[i]], method = "ML")
  ljung_p <- Box.test(residuals(fit), lag = 10, type = "Ljung-Box")$p.value
  results <- rbind(results, data.frame(
    Model = names(models)[i],
    AIC = fit$aic,
    BIC = BIC(fit),
    Ljung_Box_p = ljung_p
  ))
}
print(results)
```

```
##          Model       AIC       BIC Ljung_Box_p
## 1 ARIMA(0,1,0) -540.1976 -538.1545   0.8094629
## 2 ARIMA(1,1,0) -541.3535 -537.2674   0.7596909
## 3 ARIMA(0,1,1) -540.4525 -536.3664   0.8245419
## 4 ARIMA(1,1,1) -545.1549 -539.0258   0.8881498
## 5 ARIMA(2,1,0) -542.1543 -536.0251   0.7249147
## 6 ARIMA(2,1,1) -543.1641 -534.9919   0.8832207
## 7 ARIMA(2,1,2) -543.9988 -533.7835   0.9470827
## 8 ARIMA(1,1,2) -543.1627 -534.9904   0.8840160
## 9 ARIMA(0,1,2) -540.1068 -533.9777   0.7867186
```

```r
# If we inspect the BIC too, the one with min AIC is likely to also have the min BIC
cat("\nBest model by AIC:", results$Model[which.min(results$AIC)], "\n")
```

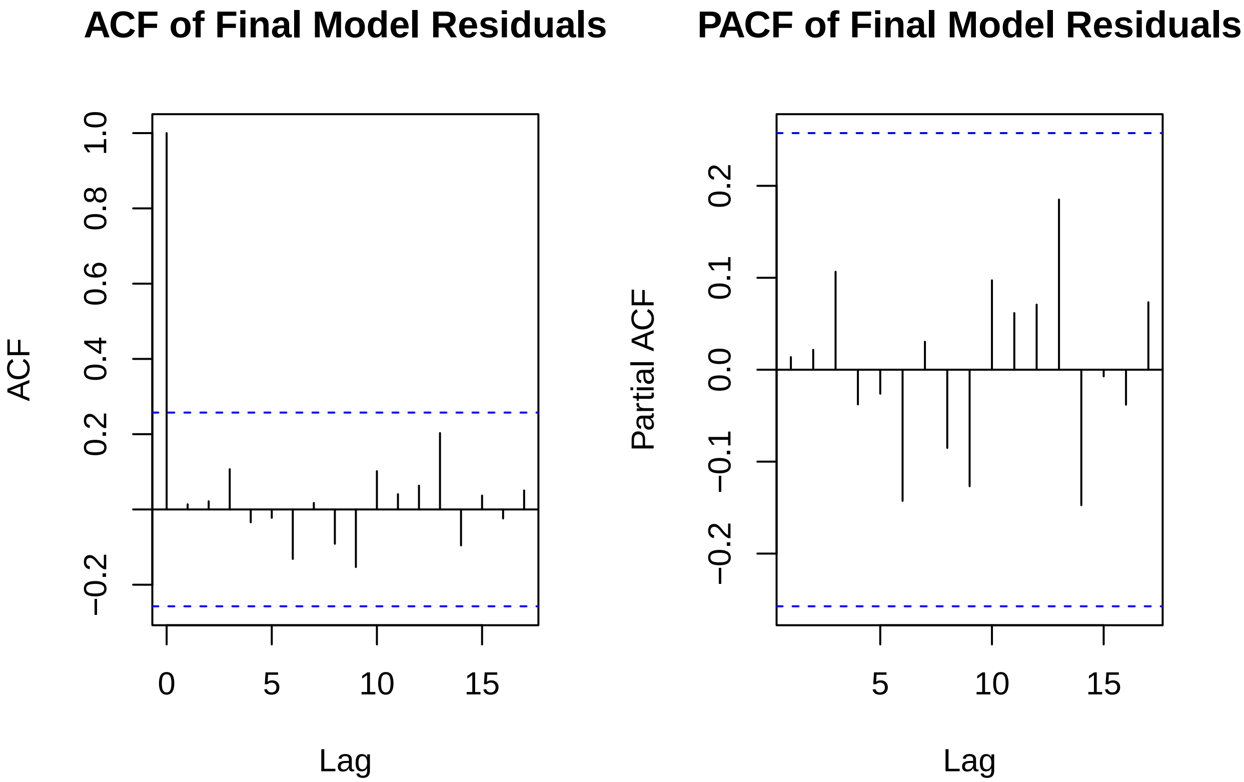
```
##
## Best model by AIC: ARIMA(1,1,1)
```

```r
# STEP 4: Final Model and Diagnostics
final_model <- Arima(boxcox_gdp_ts, order = c(1,1,1), method = "ML")
print(final_model)
```

```
## Series: boxcox_gdp_ts
## ARIMA(1,1,1)
##
## Coefficients:
##          ar1      ma1
##       0.9603  -0.8459
## s.e.  0.0808   0.1707
##
## sigma^2 = 4.215e-06:  log likelihood = 275.58
## AIC=-545.15   AICc=-544.7   BIC=-539.03
```

```r
residuals_final <- residuals(final_model)

# Residual ACF and PACF for final model
par(mfrow = c(1, 2))  # Side-by-side layout
acf(residuals_final, main = "ACF of Final Model Residuals")
```

```r
pacf(residuals_final, main = "PACF of Final Model Residuals")
```




```r
par(mfrow = c(1, 1))  # Reset layout

cat("\nDiagnostic Tests:\n")
```

```
##
## Diagnostic Tests:
```

```r
# 1. Ljung-Box test
ljung <- Box.test(residuals_final, lag = 10, type = "Ljung-Box")
cat("Ljung-Box test p-value:", round(ljung$p.value, 4),
    ifelse(ljung$p.value > 0.05, "(PASS)", "(FAIL)"), "\n")
```

```
## Ljung-Box test p-value: 0.8881 (PASS)
```

```r
# 2. Normality test
# this does not violate model assumptions, but it violates confidence interval assumptions
shapiro <- shapiro.test(residuals_final)
cat("Shapiro-Wilk test p-value:", round(shapiro$p.value, 4),
    ifelse(shapiro$p.value > 0.05, "(PASS)", "(FAIL)"), "\n")
```

```
## Shapiro-Wilk test p-value: 0.0288 (FAIL)
```

```r
# 3. ARCH test
arch <- Box.test(residuals_final^2, lag = 5, type = "Ljung-Box")
cat("ARCH test p-value:", round(arch$p.value, 4),
    ifelse(arch$p.value > 0.05, "(PASS)", "(FAIL)"), "\n")
```

```
## ARCH test p-value: 0.6157 (PASS)
```

```r
cat("\nSlight non-normality detected but acceptable for ARIMA modeling\n")
```

```
##
## Slight non-normality detected but acceptable for ARIMA modeling
```

```r
cat("Q-Q plot shows approximate normality with minor tail deviations\n\n")
```

```
## Q-Q plot shows approximate normality with minor tail deviations
```

```r
# STEP 5: Forecast with Inverse Transformation
forecast_result <- forecast(final_model, h = 3)
lambda <- 0.1
# Inverse Box-Cox transformation
forecast_original <- (lambda * forecast_result$mean + 1)^(1/lambda)
lower_original <- (lambda * forecast_result$lower + 1)^(1/lambda)
upper_original <- (lambda * forecast_result$upper + 1)^(1/lambda)
cat("1-step ahead forecast (original GDP scale):", round(forecast_original[1], 2), "million USD\n")
```

```
## 1-step ahead forecast (original GDP scale): 50.66 million USD
```

```r
cat("95% prediction interval: [", round(lower_original[1,2], 2), ",",
    round(upper_original[1,2], 2), "] million USD\n\n")
```

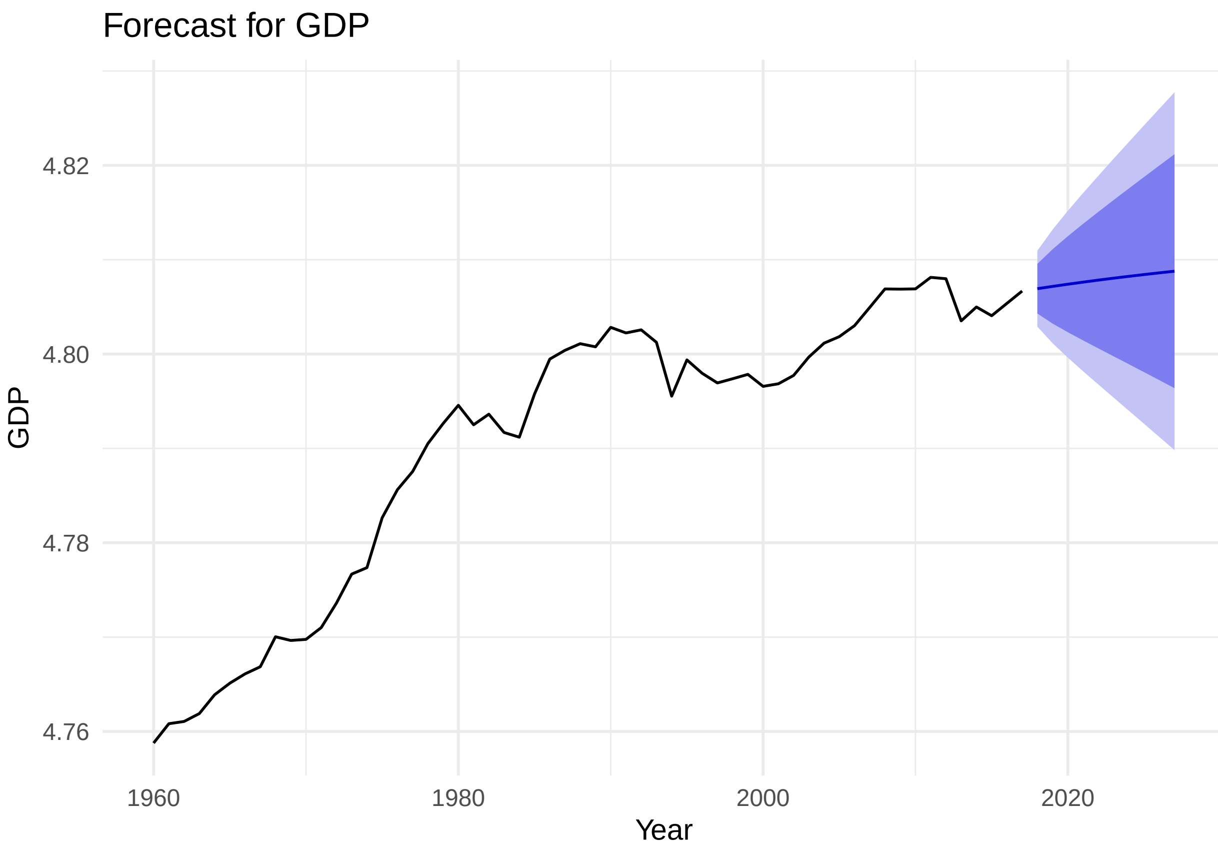
```
## 95% prediction interval: [ 50.52 , 50.8 ] million USD
```

```r
cat("FINAL MODEL: ARIMA(0,1,0) for Box-Cox transformed GDP\n")
```

```
## FINAL MODEL: ARIMA(0,1,0) for Box-Cox transformed GDP
```

Forecasting the next 10 time periods

```r
library(ggplot2)
library(forecast)

forecast_horizon <- 10
forecast_values <- forecast(final_model, h = forecast_horizon)

# Enable LaTeX rendering in plots
par(pty="m") # reset plot type if needed
options(repr.plot.width=7, repr.plot.height=5)
# Note: In R base plotting, LaTeX rendering is not native; ggplot2 with expression() or latex2exp can b
# Here we set theme with theme_minimal() and use expression for labels.

autoplot(forecast_values) +
  ggtitle(expression("Forecast for GDP")) +
  xlab(expression("Year")) +
  ylab(expression("GDP")) +
  theme_minimal()
```


Forecast for GDP
4.82
4.80
4.78
4.76
GDP
1960
1980
2000
2020
Year