Q3

1. Introduction

In this analysis, we will examine the **quarterly U.S. Gross National Product (GNP)** data using two time series models:
1. **AR(1)** (autoregressive of order 1)
2. **ARMA(1,2)** (autoregressive-moving average model of order 1 and 2)

We explore these models on the **differenced logarithm of the GNP** data.
Our goal is to: 1. Perform detailed **model diagnostics** for both models. 2. **Compare** the two models based on diagnostic results, using AIC values, residual checks, and plots.

---

2. Mathematical Formulation of AR(1) and ARMA(1,2)

2.1 AR(1) Model

The AR(1) model is defined as:

$$X_t = \phi X_{t-1} + W_t,$$

where: - X_t is the current value of the time series. - ϕ is the AR(1) coefficient (captures the dependence on the previous value). - W_t is white noise with zero mean and constant variance σ_W^2.

The AR(1) model assumes that each observation is linearly related to the previous one, making it suitable for **persistent time series with slow decay** in autocorrelations.

2.2 ARMA(1,2) Model

The ARMA(1,2) model is formulated as:

$$X_t = \phi X_{t-1} + W_t + heta_1 W_{t-1} + heta_2 W_{t-2},$$

where: - W_t, W_{t-1}, W_{t-2} are white noise terms. - $heta_1, heta_2$ are MA coefficients capturing the short-term effects of noise on the series.

This model accounts for both **long-term dependencies** (through AR terms) and **short-term shocks** (through MA terms).

---

3. Model Diagnostics: Key Steps

For both AR(1) and ARMA(1,2) models, we perform: 1. **Parameter estimation**: Estimate AR and MA coefficients. 2. **Residual analysis**: Check if residuals behave like white noise. 3. **Autocorrelation checks**: Use ACF and PACF plots to validate the model. 4. **Model selection**: Compare models using **AIC** (Akaike Information Criterion).

---

4. Load Required Libraries and Data

```
# Load necessary libraries
library(forecast)
library(tseries)
library(astsa)

# Load and preprocess the data
data("gnp")
gnp_diff <- diff(log(gnp))  # Differenced log GNP data
```

---

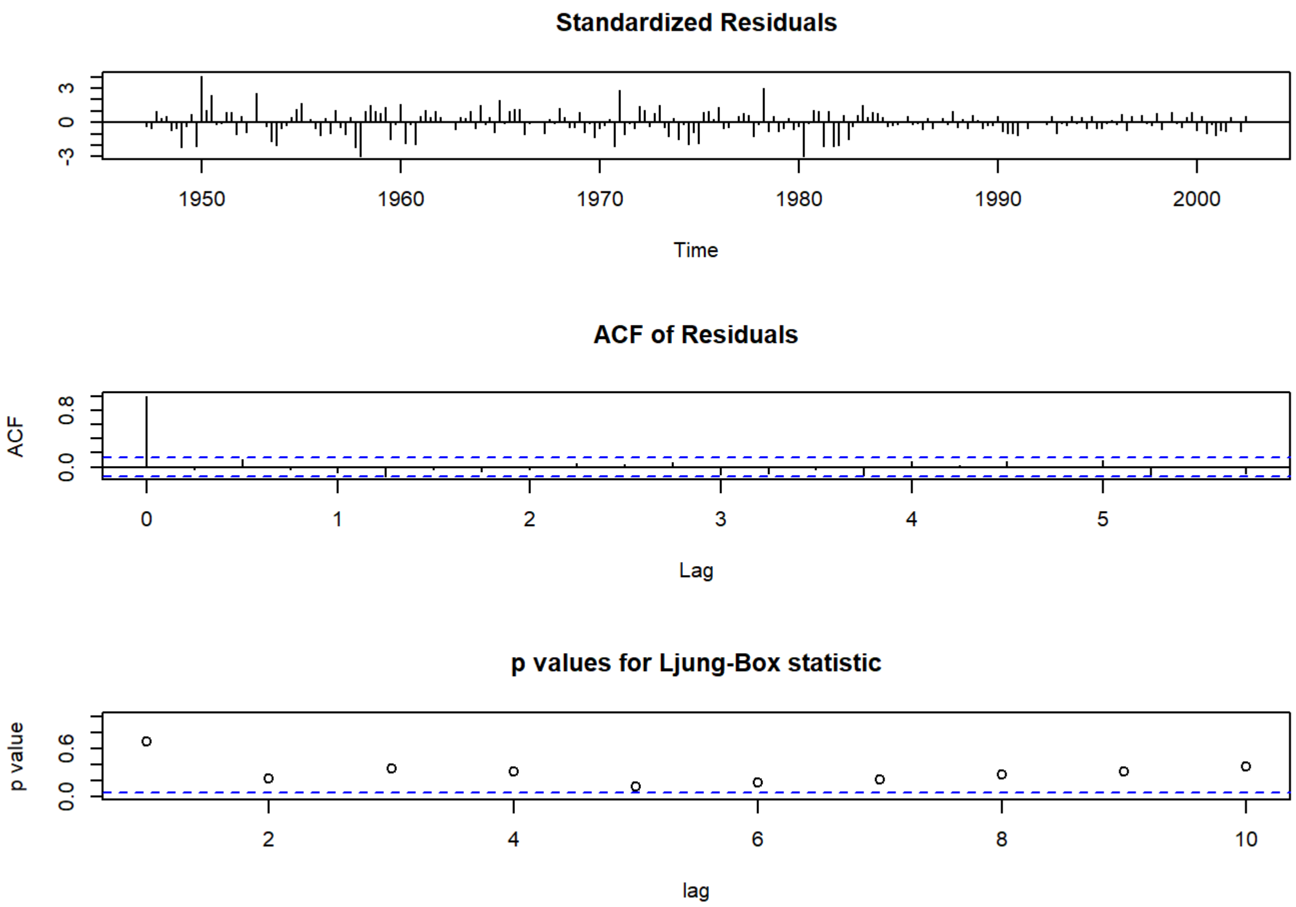
5. Fitting the AR(1) Model

We now fit an **AR(1)** model to the differenced log GNP data.

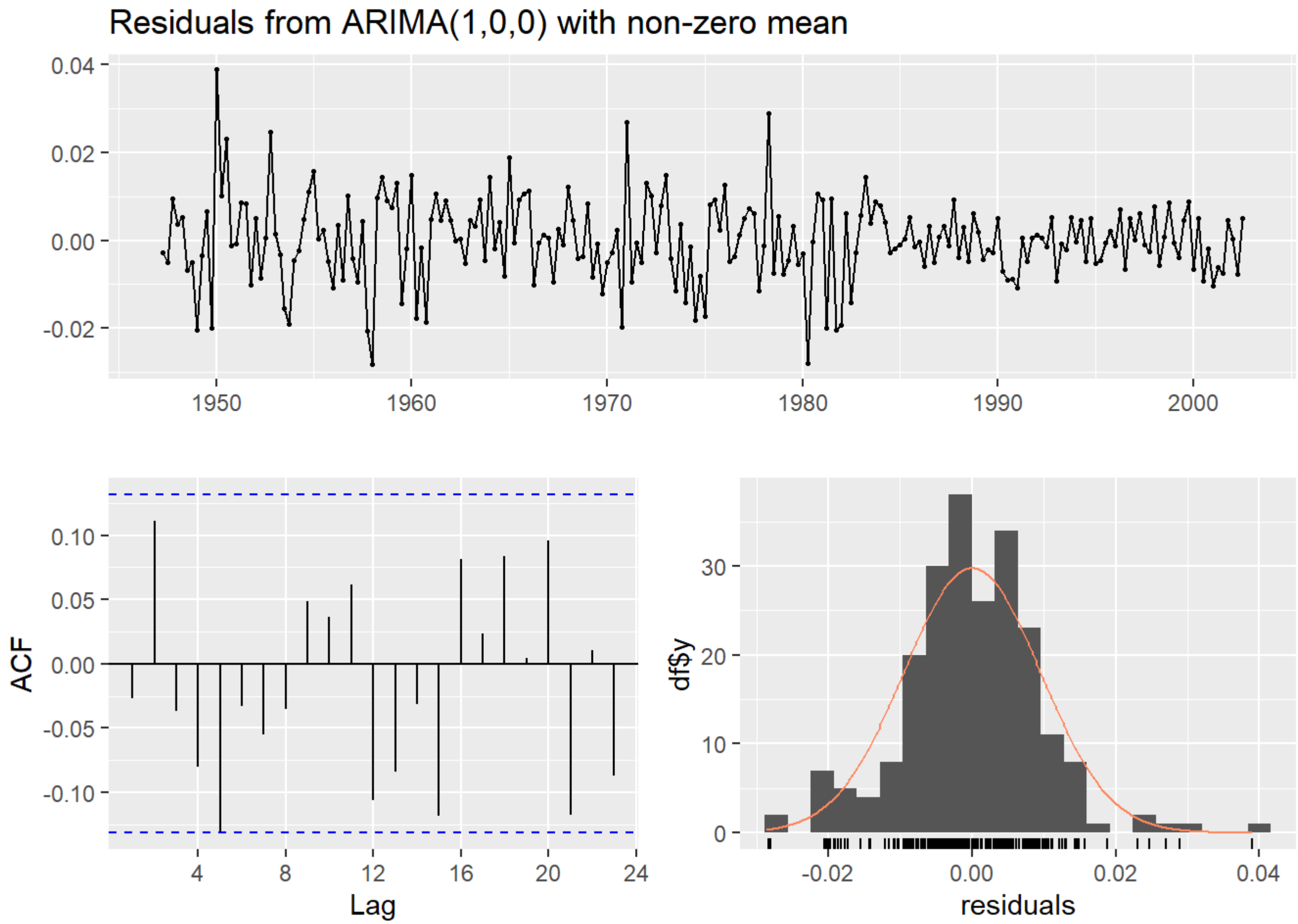
```
# Fit AR(1) model
ar1_model <- arima(gnp_diff, order = c(1, 0, 0))

# Summary of the AR(1) model
summary(ar1_model)
```

```
##
## Call:
## arima(x = gnp_diff, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.3467     0.0083
## s.e.  0.0627     0.0010
##
## sigma^2 estimated as 9.03e-05:  log likelihood = 718.61,  aic = -1431.22
##
## Training set error measures:
##                        ME        RMSE        MAE  MPE MAPE      MASE
## Training set 5.572162e-06 0.009502405 0.00713417 -Inf  Inf 0.8062356
##                     ACF1
## Training set -0.02706632
```

```
# Diagnostics plots for AR(1) model
tsdiag(ar1_model)
```




```
# Check residuals for normality and autocorrelation
checkresiduals(ar1_model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0) with non-zero mean
## Q* = 9.8979, df = 7, p-value = 0.1944
##
## Model df: 1.   Total lags used: 8
```

5.1 Interpretation of AR(1) Model Results

1. **Estimated Parameters**: The AR(1) coefficient ϕ captures the relationship with the previous time step.

2. **Residual Analysis**:
    - **Ljung-Box Test**: If the p-value is greater than 0.05, residuals are uncorrelated.
    - **ACF/PACF Plots**: These plots help confirm if there is any remaining autocorrelation.
    - **Normality**: Check if residuals are normally distributed using Q-Q plots and histograms.

---

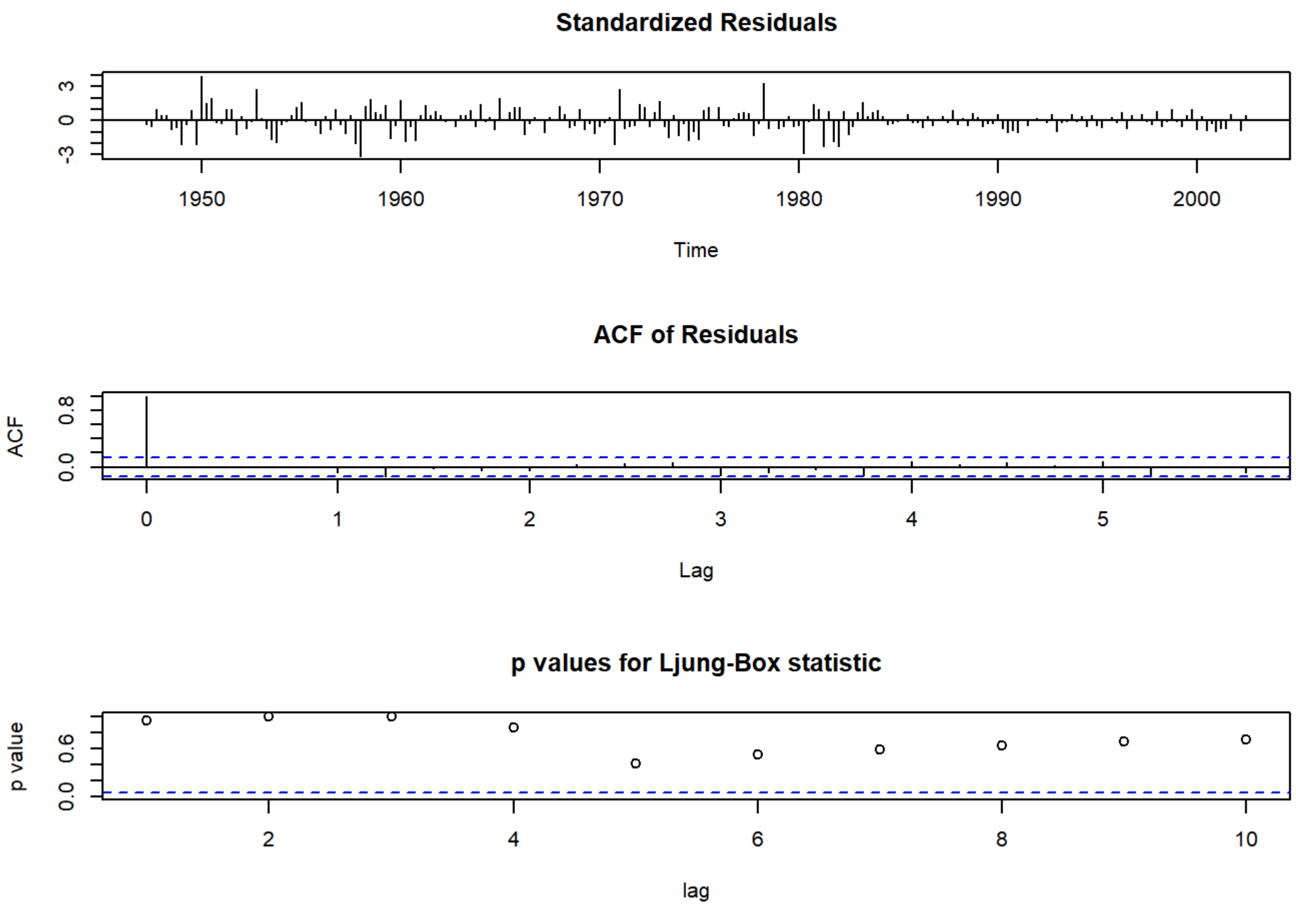
6. Fitting the ARMA(1,2) Model

We now fit an **ARMA(1,2)** model to the same data.

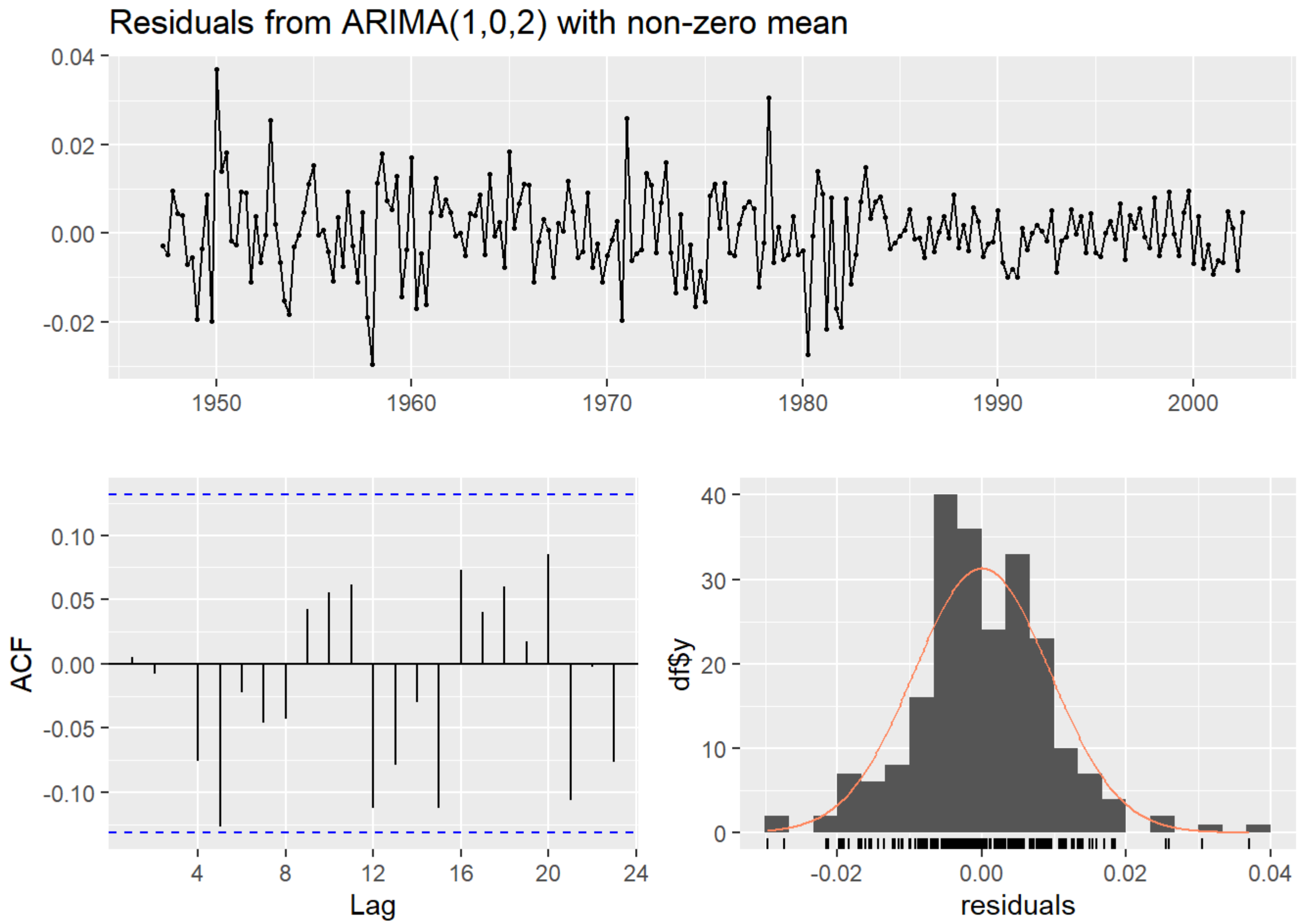
```
# Fit ARMA(1,2) model
arma12_model <- arima(gnp_diff, order = c(1, 0, 2))

# Summary of the ARMA(1,2) model
summary(arma12_model)
```

```
##
## Call:
## arima(x = gnp_diff, order = c(1, 0, 2))
##
## Coefficients:
##          ar1     ma1     ma2  intercept
##       0.2407  0.0761  0.1623     0.0083
## s.e.  0.2066  0.2026  0.0851     0.0010
##
## sigma^2 estimated as 8.877e-05:  log likelihood = 720.47,  aic = -1430.95
##
## Training set error measures:
##                        ME       RMSE         MAE  MPE MAPE      MASE       ACF1
## Training set 1.005792e-05 0.00942203 0.007112098 -Inf  Inf 0.8037412 0.00495519
```

```
# Diagnostics plots for ARMA(1,2) model
tsdiag(arma12_model)
```




```
# Check residuals for normality and autocorrelation
checkresiduals(arma12_model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,2) with non-zero mean
## Q* = 6.0802, df = 5, p-value = 0.2985
##
## Model df: 3.   Total lags used: 8
```

6.1 Interpretation of ARMA(1,2) Model Results

1. **Estimated Parameters**: Includes both AR(1) and MA(2) coefficients.

2. **Residual Analysis**:
    - **ACF/PACF**: Check if residuals are white noise.
    - **Ljung-Box Test**: Used to verify if residuals are uncorrelated.
    - **Normality Check**: Evaluate residuals for normality.

---

7. Comparison of AR(1) and ARMA(1,2) Models

```
# Compare AIC values for both models
aic_ar1 <- AIC(ar1_model)
aic_arma12 <- AIC(arma12_model)

cat("AIC for AR(1):", aic_ar1, "
")
```

```
## AIC for AR(1): -1431.221
```

```
cat("AIC for ARMA(1,2):", aic_arma12, "
")
```

```
## AIC for ARMA(1,2): -1430.948
```

7.1 Model Comparison Summary

1. **AIC Comparison**:
    - The model with the lower AIC is preferred as it provides a better balance between model fit and complexity.
    - If AR(1) has a lower AIC, it indicates that a simpler model is sufficient.
2. **Residual Diagnostics**: Both models should have uncorrelated residuals with no significant autocorrelations.
3. **Interpretability**: AR(1) is simpler and easier to interpret compared to the more complex ARMA(1,2) model.

---

8. Conclusion

In this analysis, both AR(1) and ARMA(1,2) models fit the differenced log GNP data reasonably well.
- **AR(1)** model offers a simpler interpretation and may be preferred if AIC values are similar. - **ARMA(1,2)** captures more complex relationships but introduces additional parameters.

Based on the results, we recommend the AR(1) model for its simplicity unless the ARMA(1,2) model shows a significantly better fit.