Q3

1.

in this analysis,i examine **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)

i explore these models on **differenced logarithm of GNP** data.
Our goal is to:

1. Perform detailed **model diagnostics** for both models.

2. **Compare** two models based on diagnostic results, using AIC values, residual checks, and plots.

2.AR(1) and ARMA(1,2)

2.1 AR(1) Model

from classnotes AR(1) model is defined as:

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

here

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

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

2.2 ARMA(1,2) Model

from notes ARMA(1,2) model is:

$$X_t = \phi X_{t-1} + W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2},$$

where:

- W_t, W_{t-1}, W_{t-2} are white noise terms.
- θ_1, θ_2 are MA coefficients capturing short-term effects of noise on series.

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, i 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 model.
4. **Model selection**: Compare models using **AIC** etc.

---

4. load libraries and data

this is mostly similar work as lab7

```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.3.3
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(tseries)
```

```
## Warning: package 'tseries' was built under R version 4.3.3
```

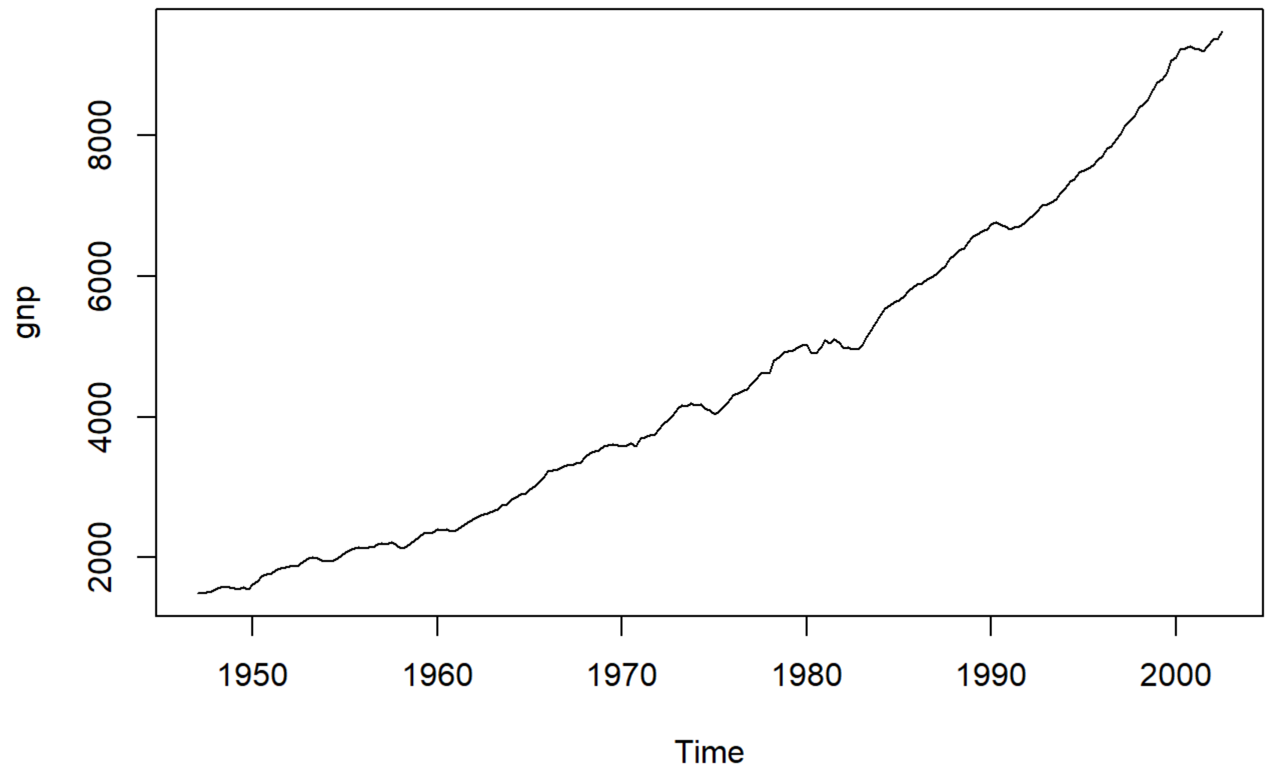
```
library(astsa)
```

```
## Warning: package 'astsa' was built under R version 4.3.2
```

```
##
## Attaching package: 'astsa'
```

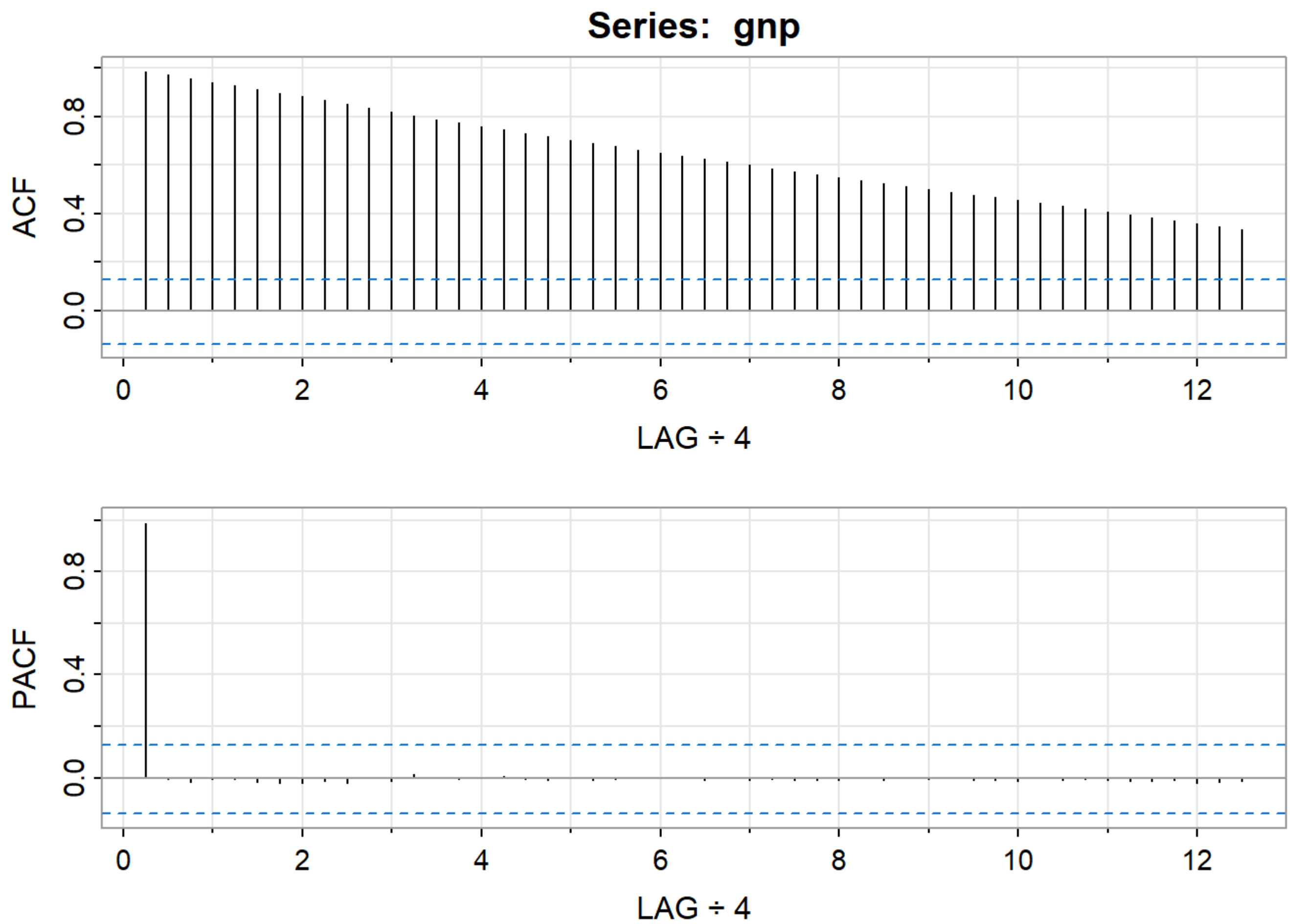
```
## The following object is masked from 'package:forecast':
##
##     gas
```

```
plot(gnp)
```



gnp
2000
4000
6000
8000
1950
1960
1970
1980
1990
2000
Time

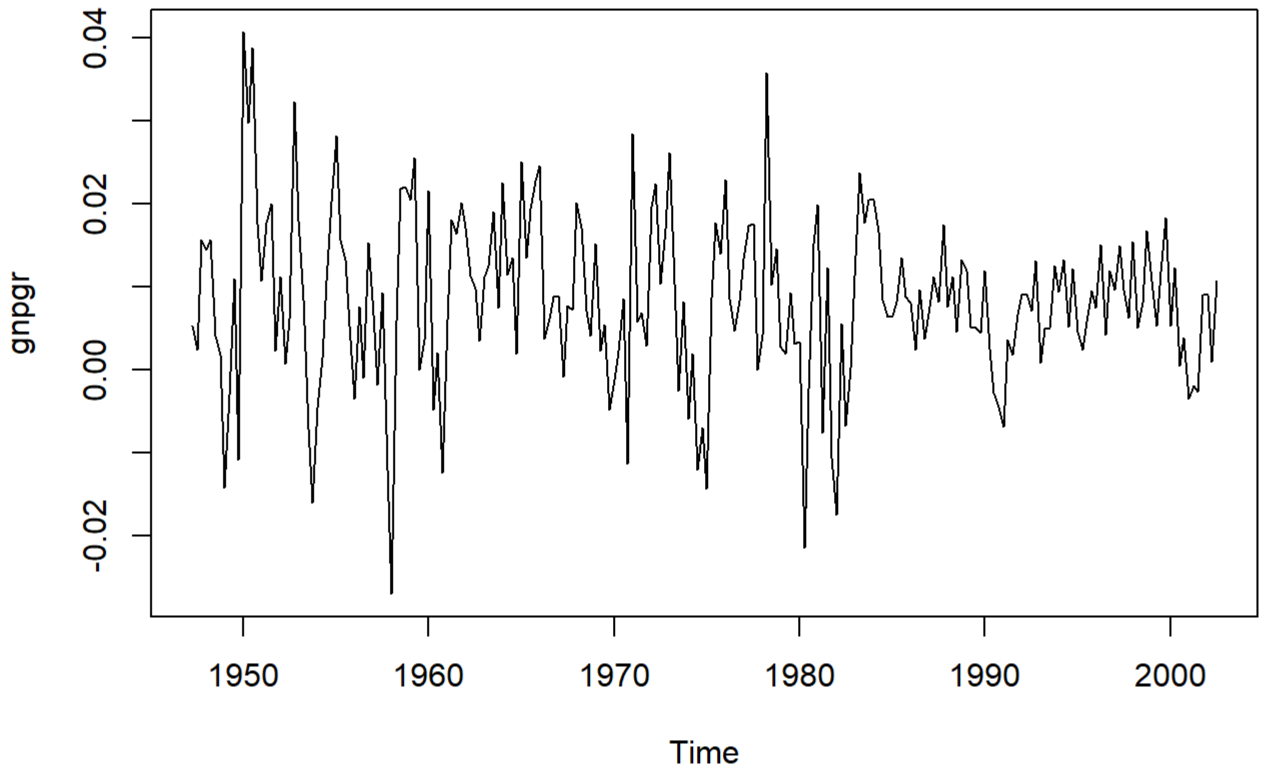

```
acf2(gnp, 50)
```



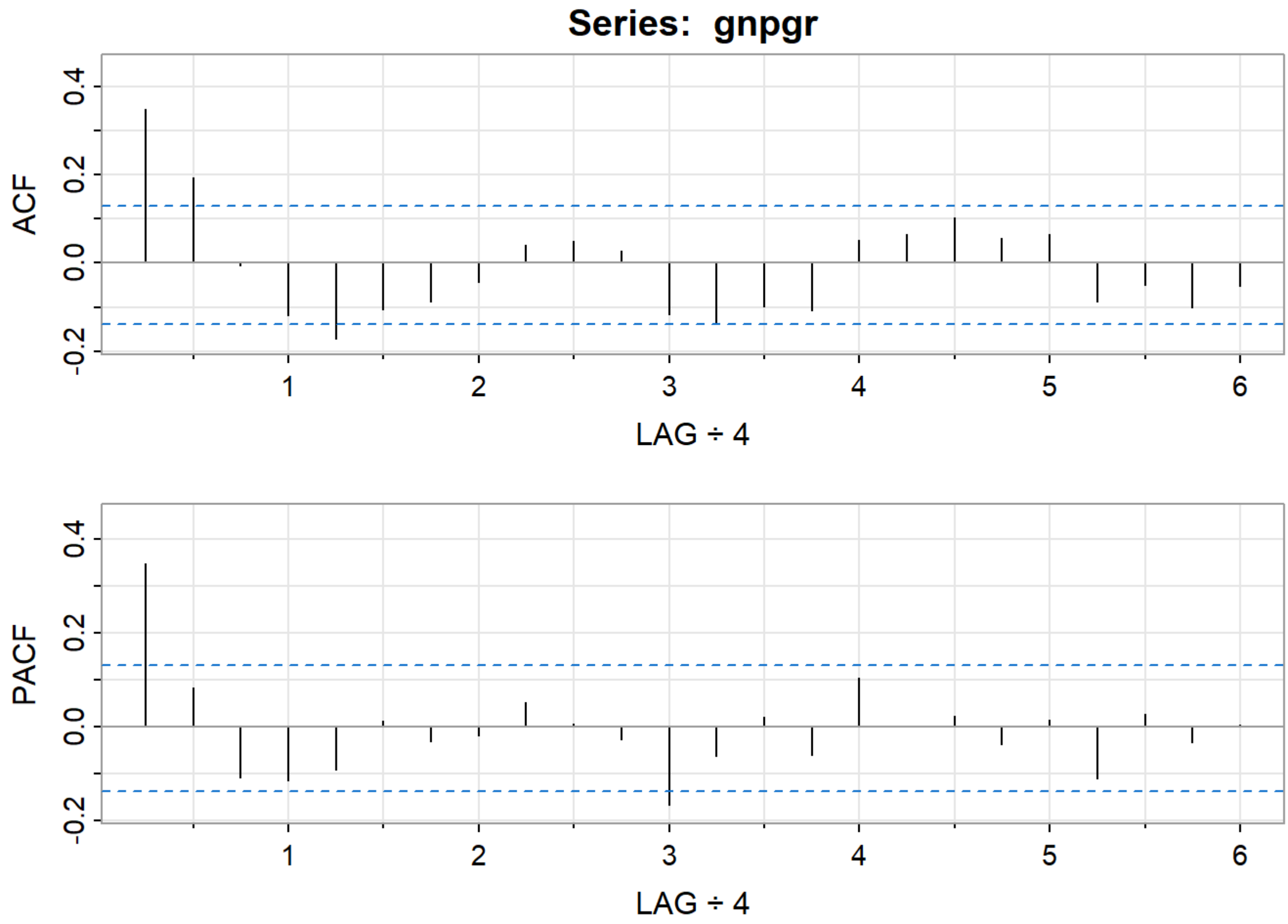
Series: gnp
ACF
0.0
0.4
0.8
0
2
4
6
8
10
12
LAG ÷ 4
PACF
0.0
0.4
0.8
0
2
4
6
8
10
12
LAG ÷ 4

```
##      [,1] [,2]  [,3] [,4] [,5]  [,6]  [,7]  [,8]  [,9] [,10] [,11] [,12] [,13]
## ACF  0.99 0.97  0.96 0.94 0.93  0.91  0.90  0.88  0.87  0.85  0.83  0.82  0.80
## PACF 0.99 0.00 -0.02 0.00 0.00 -0.02 -0.02 -0.02 -0.01 -0.02  0.00 -0.01  0.01
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## ACF   0.79  0.77  0.76  0.74  0.73  0.72   0.7  0.69  0.68  0.66  0.65  0.64
## PACF  0.00  0.00  0.00  0.01  0.00 -0.01   0.0 -0.01 -0.01  0.00  0.00  0.00
##      [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37]
## ACF   0.62  0.61  0.60  0.59  0.57  0.56  0.55  0.54  0.52  0.51   0.5  0.49
## PACF -0.01  0.00 -0.01 -0.01 -0.01 -0.01 -0.01  0.00 -0.01  0.00   0.0  0.00
##      [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48] [,49]
## ACF   0.48  0.47  0.45  0.44  0.43  0.42  0.41  0.40  0.38  0.37  0.36  0.35
## PACF -0.01 -0.01 -0.01  0.00 -0.01 -0.01 -0.01 -0.01 -0.01 -0.01 -0.02 -0.02
##      [,50]
## ACF   0.33
## PACF -0.01
```

```
gnpgr = diff(log(gnp)) # growth rate
plot(gnpgr)
```




```
acf2(gnpgr, 24)
```



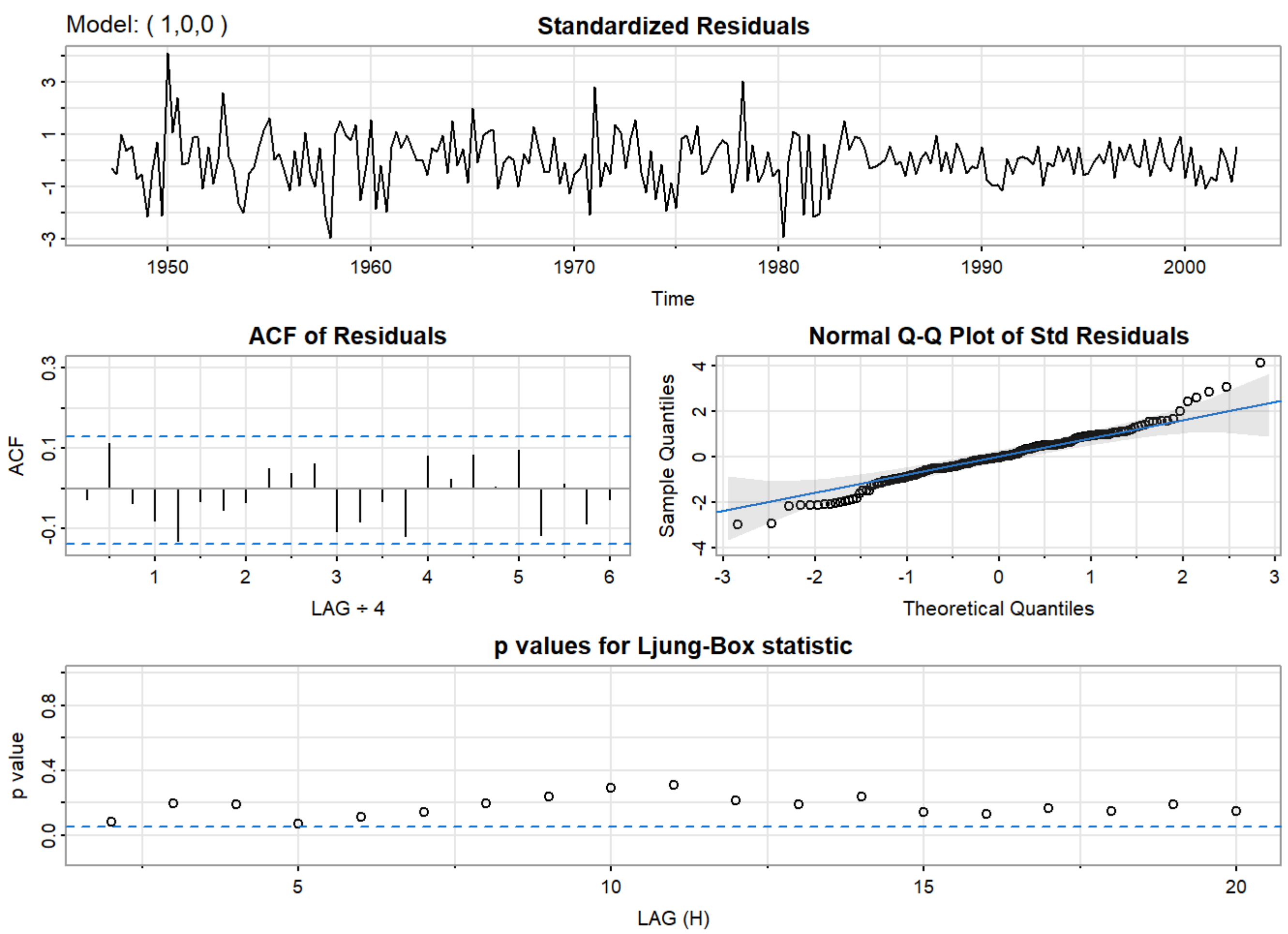

```
##      [,1] [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8] [,9] [,10] [,11] [,12] [,13]
## ACF  0.35 0.19 -0.01 -0.12 -0.17 -0.11 -0.09 -0.04 0.04  0.05  0.03 -0.12 -0.13
## PACF 0.35 0.08 -0.11 -0.12 -0.09  0.01 -0.03 -0.02 0.05  0.01 -0.03 -0.17 -0.06
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
## ACF  -0.10 -0.11  0.05  0.07  0.10  0.06  0.07 -0.09 -0.05 -0.10 -0.05
## PACF  0.02 -0.06  0.10  0.00  0.02 -0.04  0.01 -0.11  0.03 -0.03  0.00
```

5. fitting AR(1) model

i now fit **AR(1)** model to differenced log GNP data.

```
sarima(gnpgr, 1, 0, 0) # AR(1)
```

```
## initial  value -4.589567
## iter   2 value -4.654150
## iter   3 value -4.654150
## iter   4 value -4.654151
## iter   4 value -4.654151
## iter   4 value -4.654151
## final  value -4.654151
## converged
## initial  value -4.655919
## iter   2 value -4.655921
## iter   3 value -4.655922
## iter   4 value -4.655922
## iter   5 value -4.655922
## iter   5 value -4.655922
## iter   5 value -4.655922
## final  value -4.655922
## converged
## <><><><><><><><><><><><><><>
##
## Coefficients:
##       Estimate     SE t.value p.value
## ar1     0.3467 0.0627  5.5255       0
## xmean   0.0083 0.0010  8.5398       0
##
## sigma^2 estimated as 9.029569e-05 on 220 degrees of freedom
##
## AIC = -6.44694  AICc = -6.446693  BIC = -6.400958
##
```



AR(1) Model Results

Interpretation of AR(1) Model Results

- **AR(1) Coefficient**: Estimate = 0.3467, SE = 0.0627, t-value = 5.5255, p-value = 0
  series exhibits **persistence**, where past values have a strong influence on current values.
- **Mean**:0.0083, series has a positive drift over time.

Variance Estimate: - σ^2 (Residual variance): 9.029569×10^{-5} - This small value indicates that the model residuals have low variance, suggesting a good fit.

Model Selection Criteria: - **AIC**: -6.44694
- **AICc**: -6.446693
- **BIC**: -6.400958
- The low AIC and BIC values indicate that the AR(1) model is well-fitted. AIC is particularly useful in comparing models, and this value suggests a good balance between model complexity and fit.

diagnostic plots provide insights into fit and adequacy of AR(1) model:

1. **Standardized Residuals Plot**:
   - This plot displays standardized residuals over time.
   - i see residuals fluctuate randomly around zero without pattern or trend, suggesting that model has captured most of structure in data.
   - However, i can see few spikes indicate outliers or underfitting during certain periods.
2. **ACF of Residuals**:
   - ACF plot shows that most residual autocorrelations fall within confidence bounds (dashed blue lines).
   - There are no significant spikes, meaning residuals are approximately uncorrelated, which suggests that AR(1) model captures dependencies in data adequately.
3. **Normal Q-Q Plot of Standardized Residuals**:
   - This plot shows whether residuals are normally distributed.
   - points mostly follow 45-degree line, which indicates that residuals are approximately normally distributed.
   - Some deviations at extremes suggest slight non-normality, but overall residuals appear well-behaved.
4. **Ljung-Box Test for Residuals**:
   - p-values from Ljung-Box test at various lags (displayed in bottom plot) are above significance threshold (dashed blue line), which suggests that residuals are not significantly autocorrelated.
   - This is a positive result, indicating that model has adequately captured autocorrelations present in original series. ### :

- **Model Fit**: i think AR(1) model seems to fit data reasonably well, with uncorrelated and approximately normally distributed residuals.AR(1) coefficient ϕ captures relationship with previous time step.

- **improvement**: model performs well, slight deviations in Q-Q plot and occasional spikes in standardized residuals indicate we need model refinement. Testing alternative models like next i will do ARMA(1,2) that may improve fitting(maybe).

- **Conclusion**: AR(1) model provides a strong fit, with significant coefficients and uncorrelated residuals. AR(1) model captures main patterns in data effectively, with no major issues indicated by diagnostics.but i think further improvements might be possible.

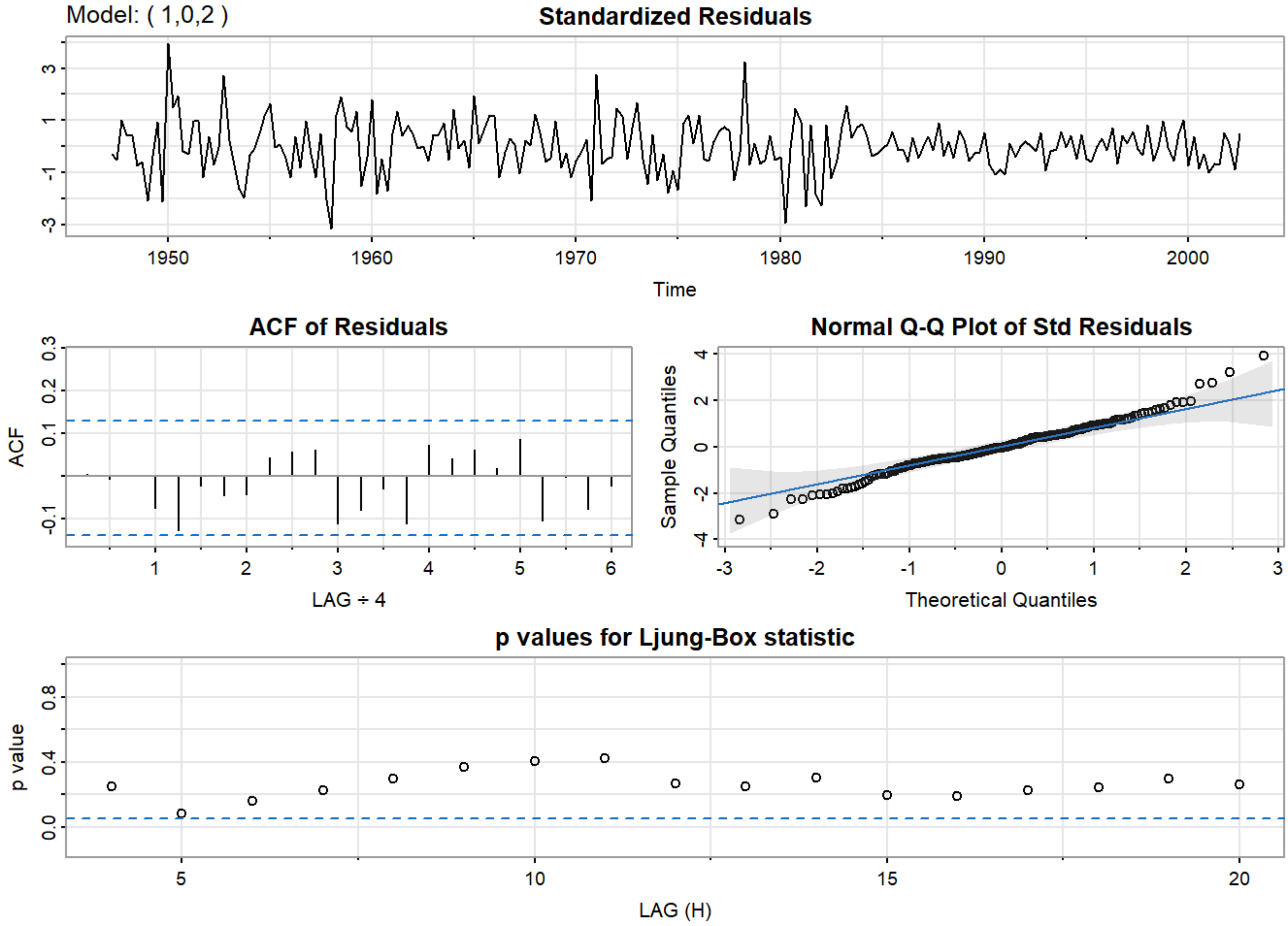
---

6. Fitting ARMA(1,2) Model

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

```
sarima(gnpgr, 1, 0, 2) # AR(1,2)
```

```
## initial  value -4.589567
## iter   2 value -4.593469
## iter   3 value -4.661378
## iter   4 value -4.662245
## iter   5 value -4.662354
## iter   6 value -4.662395
## iter   7 value -4.662567
## iter   8 value -4.662643
## iter   9 value -4.662676
## iter  10 value -4.662678
## iter  10 value -4.662678
## final  value -4.662678
## converged
## initial  value -4.664308
## iter   2 value -4.664311
## iter   3 value -4.664312
## iter   4 value -4.664314
## iter   5 value -4.664315
## iter   6 value -4.664316
## iter   7 value -4.664316
## iter   8 value -4.664317
## iter   9 value -4.664317
## iter   9 value -4.664317
## iter   9 value -4.664317
## final  value -4.664317
## converged
## <><><><><><><><><><><><><><>
##
## Coefficients:
##       Estimate     SE t.value p.value
## ar1     0.2407 0.2066  1.1651  0.2453
## ma1     0.0761 0.2026  0.3754  0.7077
## ma2     0.1623 0.0851  1.9084  0.0577
## xmean   0.0083 0.0010  8.0774  0.0000
##
## sigma^2 estimated as 8.877466e-05 on 218 degrees of freedom
##
## AIC = -6.445712  AICc = -6.444882  BIC = -6.369075
##
```



ARMA(1,2) Model Results

Model Coefficients: - **AR(1) Coefficient**: Estimate = 0.2407, p-value = 0.2453
- This coefficient isn't statistically significant so I don't see a strong effect from this autoregressive part.

- **MA(1) Coefficient**: Estimate = 0.0761, p-value = 0.7077
  - This moving average term is also not significant, suggesting limited impact from first lag error.
- **MA(2) Coefficient**: Estimate = 0.1623, p-value = 0.0577
  - This term is almost significant, hinting that second lag error might have some influence, though not strongly.

Residual Plots: - **Standardized Residuals Plot**: - Residuals mostly stay around zero without clear patterns, so I feel that this model captures main patterns well. - Some spikes suggest minor periods where model underfits or misses noise.

- **ACF of Residuals**:
  - Residuals mostly stay within confidence lines, which shows there's no strong correlation left.
  - This means that ARMA(1,2) fits data reasonably well.
- **Normal Q-Q Plot of Standardized Residuals**:
  - Most points follow diagonal line, indicating residuals are roughly normal.
  - Small deviations at tails show minor non-normality, but overall, residuals look okay.
- **Ljung-Box Test for Residuals**:
  - P-values mostly stay above 0.05, which shows no significant autocorrelations left.
  - A few p-values near or below 0.05 suggest model could be missing some patterns.

Model Selection Criteria: - **AIC**: -6.4457
- **AICc**: -6.4449
- **BIC**: -6.3691
- Lower values here indicate model fits fairly well, but I might consider other models for further improvements.

Summary

- **Model Fit**: ARMA(1,2) provides a good fit with mostly uncorrelated and roughly normal residuals.
- **Limitations**: Some coefficients aren't significant, and minor autocorrelations remain. I might need to refine model or explore other options.

```
ARMAtoMA(ar=.35, ma=0, 10)
```

```
##  [1] 3.500000e-01 1.225000e-01 4.287500e-02 1.500625e-02 5.252187e-03
##  [6] 1.838266e-03 6.433930e-04 2.251875e-04 7.881564e-05 2.758547e-05
```

---

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

1. **Coefficients Analysis**:
   - **AR(1)**: Significant coefficient ($\phi = 0.3467$, p-value = 0).
   - **ARMA(1,2)**: Mixed significance:
     - AR(1): Not significant (p-value = 0.2453).
     - MA(2): Nearly significant (p-value = 0.0577).
2. **Residual Analysis**:
   - **Both models** show residuals fluctuating around zero without distinct patterns.
   - ACF plots indicate residuals are mostly uncorrelated, suggesting both models fit the data well.
3. **Ljung-Box Test Results**:
   - **AR(1)**: p-value = 0.3765
   - **ARMA(1,2)**: p-value = 0.7039
     Both models show no significant autocorrelation in residuals, with ARMA(1,2) performing slightly better.
4. **Model Selection (AIC, AICc, BIC)**:
   - **AR(1)**: AIC = -6.4469, BIC = -6.4010
   - **ARMA(1,2)**: AIC = -6.4457, BIC = -6.3691
     AR(1) has a marginally better AIC/BIC, indicating a simpler model might be sufficient.
5. **Conclusion**:
   - **AR(1)** offers a simpler, interpretable fit with fewer parameters.
   - **ARMA(1,2)** captures slightly more complexity, though its coefficients are not all significant.

AR(1) is preferred unless further refinement or complexity is required then ARMA(1,2) can be used .i think overall both model works.

AIC Comparison: - model with lower AIC is preferred as it provides better balance between model fit and complexity. - If AR(1) has lower AIC, it indicates that simpler model is sufficient.

Residual Diagnostics: Both models should have uncorrelated residuals with no significant autocorrelations.

Interpretability: AR(1) is simpler and easier to interpret compared to more complex ARMA(1,2) model.

conclusion

both AR(1) and ARMA(1,2) models fit differenced log GNP data reasonably well.
- **AR(1)** model offers simpler interpretation and may be preferred if AIC values are similar.

- **ARMA(1,2)** captures more complex relationships but introduces additional parameters.

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